



**Universidad de Santiago de Compostela**

Facultad de Medicina y Odontología

Departamento de Medicina

**FACTORES GENÉTICOS IMPLICADOS EN EL  
DESARROLLO DEL LUPUS ERITEMATOSO  
SISTÉMICO:**

**Estudio de polimorfismos funcionales.**

**Tesis Doctoral**

**Elisa Alonso Pérez**

**2012**

**Universidad de Santiago de Compostela**

Facultad de Medicina y Odontología

Departamento de Medicina

**Hospital Clínico Universitario de Santiago de Compostela**

Instituto de Investigación Sanitaria de Santiago

Laboratorio de Investigación 10

**FACTORES GENÉTICOS IMPLICADOS EN EL  
DESARROLLO DEL LUPUS ERITEMATOSO  
SISTÉMICO:**

**Estudio de polimorfismos funcionales.**

**Tesis Doctoral**

**Elisa Alonso Pérez**

**2012**

El Dr. Antonio González Martínez-Pedrayo, investigador principal del Laboratorio de Investigación 10 del Instituto de Investigación Sanitaria del Complejo Hospitalario Universitario de Santiago de Compostela, y el Prof. Dr. Juan Jesús Gómez-Reino Carnota, Jefe del Servicio de Reumatología del Hospital Clínico Universitario de Santiago de Compostela y Catedrático de Medicina del Departamento de Medicina de la Universidad de Santiago de Compostela,

**CERTIFICAN QUE:**

El presente trabajo que lleva por título: *Factores genéticos implicados en el desarrollo del lupus eritematoso sistémico: Estudio de polimorfismos funcionales*, realizado por Elisa Alonso Pérez en el Departamento de Medicina de la Universidad de Santiago de Compostela bajo nuestra dirección, ha sido revisado y está en disposición de ser presentado para optar al grado de Doctora en Biología.

Fdo: Dr. Antonio González Martínez-Pedrayo    Fdo: Prof. Dr. Juan J. Gómez-Reino Carnota

## ÍNDICE.

---



## ÍNDICE:

<b>ABREVIATURAS .....</b>	<b>1</b>
<b>INTRODUCCIÓN.....</b>	<b>6</b>
1. Lupus Eritematoso Sistémico .....	6
1.1. Clínica del LES .....	7
1.2. Patogenia del LES .....	9
1.3. Papel de interferón de tipo I (IFN-I) en la patogenia del LES .....	10
1.4. Etiología del LES .....	13
2. Estudio de la genética del LES.....	13
2.1. Estudios de ligamiento .....	14
2.2. Estudios de asociación.....	15
2.3. Estudios de asociación de genoma completo .....	16
2.3.1 GWAS de LES en Europeos .....	19
2.3.2 GWAS de LES en Asiáticos.....	21
2.3.3. La heredabilidad no explicada del LES.....	24
3. Factores genéticos del LES .....	25
3.1. Región HLA .....	25
3.2. Factores del Complemento .....	27
3.3. <i>TREX1</i> .....	28
3.4. FCGRs .....	28
3.5. <i>TYK2</i> .....	30
3.6. <i>STAT4</i> .....	31
3.7. <i>PTPN22</i> .....	33
3.8. <i>PDCD1</i> .....	34
3.9. <i>TNFAIP3</i> y la región intergénica <i>6q23.TNIP1</i> .....	35
3.10. <i>ITGAM</i> .....	37
3.11. <i>C8orf13-BLK</i> .....	38

3.12. <i>BANK1-LYN</i> .....	38
3.13. <i>IRAK1-MECP2</i> .....	40
3.14. <i>TNFSF4</i> .....	41
3.15. <i>PHRF1-IRF7</i> .....	42
3.16. <i>ATG5-PRDM1-ETS1</i> .....	43
3.17. <i>LY9</i> .....	43
4. Nuevos genes asociados al LES .....	45
5. Factor 5 regulador del IFN ( <i>IRF5</i> ) .....	46
5.1. Asociación genética de <i>IRF5</i> con la susceptibilidad al LES .....	49
5.1.1. SNP rs2004640: <i>splice donor site</i> en el exón 1B .....	50
5.1.2. SNP rs10954213: señal de poly A.....	51
5.1.3. Inserción/delección presente en el exón 6: región rica en prolina .....	53
5.1.4 Inserción/delección CGGGG: unión del factor de transcripción SP1 .....	54
6. Estudio de la expresión génica: <i>microarrays</i> de expresión y secuenciación del ARN (RNA-seq) .....	55
7. Heterogeneidad del LES.....	57
7.1. Heterogeneidad clínica .....	57
7.2. Heterogeneidad genética .....	58
<b>HIPÓTESIS .....</b>	<b>62</b>
<b>OBJETIVOS .....</b>	<b>64</b>
<b>PROCEDIMIENTO EXPERIMENTAL Y RESULTADOS .....</b>	<b>65</b>
1. Esquema general.....	66
2. Heterogeneidad clínica del LES: Publicación 1 .....	69
3. Heterogeneidad genética del LES: Publicación 2 .....	77
4. Estudio de la variabilidad genética de <i>IRF5</i> en relación al LES. Publicación 3 .....	97
5. Estudio de la regulación de la expresión génica de <i>IRF5</i> mediante datos de expresión procedentes de estudios con <i>microarrays</i> : Publicación 4 .....	112
6. Mapa de LD de <i>IRF5</i> : búsqueda de polimorfismos funcionales .....	139
6.1. Material y Métodos .....	140

6.2. Resultados .....	158
<b>DISCUSIÓN.....</b>	<b>188</b>
<b>CONCLUSIONES .....</b>	<b>211</b>
<b>BIBLIOGRAFÍA .....</b>	<b>213</b>

## Abreviaturas

3'UTR	Región 3' no traducida (del inglés, <i>3' untranslated region</i> )
5'UTR	Región 5' no traducida (del inglés, <i>5' untranslated region</i> )
ACR	Colegio Americano de Reumatología (del inglés, <i>American College of Rheumatology</i> )
ADN	Ácido desoxirribonucleico
ADNdc	Ácido desoxirribonucleico doble cadena
ANA	Anticuerpo antinuclear (del inglés, <i>antinuclear antibody</i> )
anti-ADNdc	Anticuerpo anti-ADN de doble cadena
aPL	Anticuerpo antifosfolípido
anti-RBP	Anticuerpos frente proteínas de unión al ARN (del inglés, <i>RNA binding protein</i> )
anti-RP	Anticuerpos frente fosfoproteínas ribosomales (del inglés, <i>ribosomal phosphotrein</i> )
AR	Artritis reumatoide
ARN	Ácido ribonucleico
ARNcs	Ácido ribonucleico de cadena simple
ARNdc	Ácido ribonucleico de doble cadena
ARNr	Ácido ribonucleico ribosomal
BANK1	Proteína estructural de células B con repeticiones de anquirina (del inglés, <i>B-cell scaffold protein with ankyrin repeats</i> )
BCR	Receptor de células B (del inglés, <i>B-cell receptor</i> )
BLK	Quinasa de linfocito B (del ingles, <i>B-lymphocyte kinase</i> )
C3bi	Producto de activación del complemento 3
C4A	Componente del complemento 4A
C4B	Componente del complemento 4B
CBF	<i>Core binding factor</i>
CFB	Factor de complemento B
CNV	Variantes en el número de copias (del inglés, <i>copy number variation</i> )
CPA	Célula presentadora de antígeno
CREB1	Proteína 1 de unión al elemento de respuesta al cAMP (del ingles, <i>cAMP responsive element binding protein 1</i> )
CSK	Tirosín-quinasa Src C-terminal (del inglés, <i>C-terminal Src tyrosine kinase</i> )
DC	Célula dendrítica (del inglés, <i>dendritic cell</i> )
EMSA	Ensayo de cambio de movilidad electroforética (del inglés, <i>electrophoretic mobility shift assay</i> )
ENCODE	<i>ENCyclopedia Of DNA Elements</i>
ETS1	<i>Protein C-ets-1</i>
FCGR	Receptor de baja afinidad de la fracción cristizable de la inmunoglobulina G (del inglés, <i>Fc fragment of IgG low affinity receptor</i> )

GSMA	Meta-análisis de búsqueda genómica (del inglés, <i>genome-search meta-analysis</i> )
GWAS	Estudio de asociación de genoma completo (del inglés, <i>genome wide association study</i> )
HIC2-UBE2L3	<i>Hypermethylated in cancer 2-ubiquitin-conjugating enzyme E2L 3</i>
HLA	Antígeno leucocitario humano (del inglés, <i>human leukocyte antigen</i> )
IC	Inmuno-complejos
ICA1	<i>Islet cell autoantigen 1</i>
ICAM1	Molécula 1 de adhesión intracelular (del inglés, <i>intracellular adhesion molecule 1</i> )
IFN	Interferón
IFN-I	Interferón de tipo I
IFNAR	Receptor de interferón $\alpha$ (del inglés, <i>IFN<math>\alpha</math> receptor</i> )
IFNGR2	Receptor 2 del IFN $\gamma$
Ig	Inmunoglobulina
IKZF1	<i>Ikaros family zinc finger protein 1</i>
IL	Interleucina
IL1R	Receptor de interleucina 1 (del inglés, <i>IL1 receptor</i> )
indel	Polimorfismo del tipo inserción/delección
IP <sub>3</sub> R	Receptor de inositol 1,4,5-trifosfato (del inglés, <i>inositol 1,4,5-triphosphate receptor</i> )
IRAK1	Quinasa 1 asociada a receptor de interleucina 1 (del inglés, <i>interleukin 1 receptor associated kinase 1</i> )
IRF	Factor regulador del interferón (del inglés, <i>interferon regulatory factor</i> )
ITG	Integrina
ITGAL	Integrina $\alpha$ -L
ITGAM	Integrina $\alpha$ -M
ITGAX	Integrina $\alpha$ -X
JAK	<i>Janus</i> kinasa
JAZF1	<i>JAZF zinc finger 1</i>
LCN2	Lipocalina asociada a gelatinasas de neutrófilos 2
LD	Desequilibrio de ligamiento (del inglés, <i>linkage disequilibrium</i> )
LES	Lupus eritematoso sistémico
LPS	Lipolisacáridos
LRRC18	<i>Leucine rich repeat containing 18</i>
LYN	<i>Yamaguchi sarcoma viral oncogene homolog</i>
LYP	Fosfatasa de tirosina linfóide (del inglés, <i>lymphoid-specific tyrosine phosphatase</i> )
mDC	Célula dendrítica mielóide (del inglés, <i>myeloid dendritic cell</i> )
MECP2	Proteína 2 de unión a CpG metilado (del inglés, <i>methyl-CpG-binding protein 2</i> )
MHC	Complejo mayor de histocompatibilidad (del inglés, <i>major histocompatibility complex</i> )
MICB	Polipéptido relacionado con la secuencia B, MHC clase I

MMP14	Metalo-proteasa de la matriz extracelular 14
MSH5	<i>MutS</i> homólogo 5
NCF2	<i>Neutrophil cytosol factor 2</i>
NF $\kappa$ B	Factor nuclear $\kappa$ B (del inglés, <i>nuclear factor <math>\kappa</math>B</i> )
NK	Células “asesinas naturales” (del inglés, <i>natural killer</i> )
NOD2	<i>nucleotide-binding oligomerization domain containing 2</i>
nsSNP	SNP no sinónimo
OR	<i>Odds ratio</i>
OR2H2	Receptor olfatorio 2H2
PARP	Poli-ADP-ribosa polimerasa
PBMCs	Células mononucleares de sangre periférica (del inglés, <i>peripheal blood mononuclear cells</i> )
pDC	Célula dendrítica plasmacitoide (del inglés, <i>plasmacytoid dendritic cell</i> )
PDCD1	Gen de muerte celular programada 1 (del inglés, <i>programmed cell death 1</i> )
PHRF1	<i>PHD and ring finger domain-containing protein 1</i>
PRDM1-ATG5	<i>PR domain containing 1, with ZNF domain-APG5 autophagy 5-like</i>
PRF1	Perforina 1
PTPN22	Fosfatasa de tirosina tipo 22 no-receptor (del inglés, <i>protein tyrosin phosphatase non-receptor type 22</i> )
PXK	Quinasa de serina-treonina con dominio homólogo a <i>Phox</i> (del inglés, <i>Phox homology domain-containing serine-threonine kinase</i> )
RASGRP3	<i>Ras guanyl-releasing protein 3</i>
RUNX1	<i>Runt-related transcription factor 1</i>
SCUBE1	<i>Signal peptide CUB domain, EGF-like 1</i>
SKIV2L	Helicasa SKIV2L
SLC15A4	<i>Solute carrier family 15</i>
SLEGEN	Consorcio Internacional de Genética del Lupus Eritematoso Sistémico (del inglés, <i>Systemic Lupus Erythematosus Genetics</i> )
SNP	Polimorfismo de un sólo nucleótido (del inglés, <i>single nucleotide polymorphism</i> )
STAT	Transductor de señal y activador de la transcripción (del inglés, <i>signal transducer and activator of transcription</i> )
SV40	<i>Simian virus 40</i>
T1D	Diabetes tipo 1 (del inglés, <i>type 1 diabetes</i> )
tag SNP	SNP marcador
TCR	Receptor de células T (del inglés, <i>T-cell receptor</i> )
TDT	Análisis de desequilibrio de transmisión (del inglés, <i>transmission disequilibrium test</i> )
Th	Linfocito T colaborador (del inglés, <i>T helper</i> )
TLR	Receptor tipo Toll (del inglés, <i>Toll-like receptor</i> )
TNF	Factor de necrosis tumoral (del inglés, <i>tumor necrosis factor</i> )

TNFAIP3	Proteína 3 inducida por el factor de necrosis tumoral $\alpha$ (del inglés, <i>tumor necrosis factor <math>\alpha</math>-induced protein 3</i> )
TNFR	Receptor del TNF
TNFSF4/OX40L	Miembro 4 de la superfamilia de ligandos del factor de necrosis tumoral (del inglés, <i>tumor necrosis factor (ligand) superfamily, member 4</i> )
TNIP1	Proteína de interacción con TNFAIP3 (del inglés, <i>TNFAIP3-interacting protein 1</i> )
TNPO3	<i>Transportin-3</i>
TRAF	Factor asociado al receptor del factor de necrosis tumoral (del inglés, <i>tumor necrosis factor receptor-associated factor</i> )
Treg	Linfocito T regulador
TREX1	Exonucleasa 3' reparadora 1 (del inglés, <i>three prime repair exonuclease 1</i> )
TYK2	Tirosín-quinasa 2
UHRF1BP1	Proteína 1 de unión a ICBP90 (del inglés, <i>ICBP90 binding protein 1</i> )
WDFY4	<i>WD repeat- and FYVE domain-containing protein 4</i>
XKR6	<i>Kell blood group complex subunit-related family, member 6</i>
$\lambda$	Factor de inflación
$\lambda_s$	Riesgo relativo entre hermanos

## **INTRODUCCIÓN.**

---



## **1. Lupus Eritematoso Sistémico.**

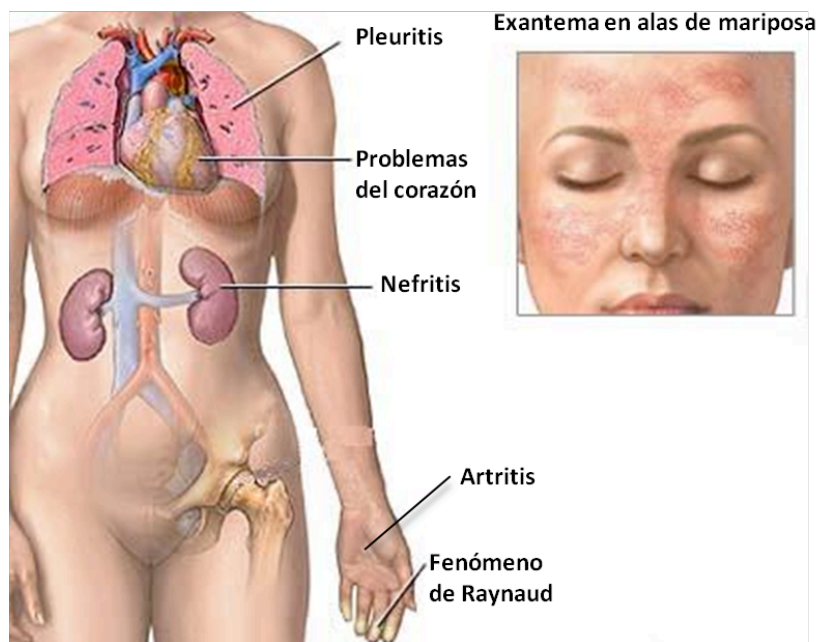
El lupus eritematoso sistémico (LES; [MIM] 152700) es una enfermedad autoinmune, multi-sistémica y crónica caracterizada por la presencia de autoanticuerpos patogénicos frente a antígenos intracelulares ubicuos. La producción de autoanticuerpos es debida a alteraciones en el sistema inmune innato y adquirido que llevan a la pérdida de la tolerancia frente a antígenos propios.

El LES se caracteriza por afectar a múltiples sistemas orgánicos. Entre sus manifestaciones clínicas son frecuentes los síntomas inespecíficos como malestar general, fatiga, fiebre y pérdida de peso. Además, la mayoría de los pacientes con LES presentan lesiones en la piel, articulaciones, riñones y membranas serosas. Destaca por sus repercusiones e incidencia la nefropatía, la cual afecta a un tercio de los pacientes con LES. El LES también se asocia con una gran variedad de trastornos del sistema nervioso, gastrointestinales y cardiovasculares. La evolución de la enfermedad es muy variable ya que puede ser crónica, remitente o, lo más frecuente, recidivante con exacerbaciones y remisiones (**Figura 1**).

El LES afecta principalmente a mujeres en edad fértil con un pico de incidencia entre los 15 y 40 años de edad. La relación mujer:hombre es 9:1, pero en los años previos y posteriores a la edad fértil de la mujer, la relación mujer:hombre disminuye a 2:1 - 3:1<sup>1</sup>.

La prevalencia global del LES es 1:1000, pero varía considerablemente dependiendo de la población<sup>2</sup>. El LES presenta una prevalencia del 0.02-0.06% en población europea, mientras que en poblaciones asiáticas, afro-americanas e hispano-americanas la prevalencia es 3 o 4 veces mayor<sup>3, 4</sup>. Sin embargo, incluso dentro de la población europea existen diferencias en prevalencia. Por ejemplo, la prevalencia de la enfermedad en Inglaterra y Gales es de 12,5 casos por 100000 mujeres de todas las edades y 17,7 casos por 100000 mujeres entre los 15-64 años de edad<sup>5</sup>. Otros dos estudios realizados en Nottingham revelaron prevalencias de 24,6 casos por 100000 personas<sup>6</sup> y de 27,7 casos por 100000 personas<sup>3</sup>. La prevalencia más elevada encontrada en Europa es de 36,3 casos por 100000 personas en Suecia<sup>7</sup>.

Estas diferencias en la prevalencia del LES se atribuyen a múltiples factores: genéticos, socio-económicos, ambientales, inmunológicos e incluso problemas de diagnóstico.



**Figura 1:** Esquema resumen de las manifestaciones clínicas más comunes del LES.

### **1.1. Clínica del LES:**

El diagnóstico del LES resulta complicado debido a la gran heterogeneidad de su presentación clínica. Según el Colegio Americano de Reumatología (ACR), un paciente debe cumplir al menos 4 de una lista de 11 criterios para la clasificación como paciente con LES<sup>8-10</sup> (**Tabla 1**). Las manifestaciones clínicas se pueden dividir en leves, como la fotosensibilidad, úlceras orales y artritis; y graves, como las alteraciones renales, convulsiones, psicosis y anormalidades hematológicas severas.

En el LES se produce una gran variedad de autoanticuerpos de isotipo IgG que tienen alta afinidad por el antígeno y especificidad restringida. Se han descrito algo más de 100 autoanticuerpos diferentes en los pacientes con LES, entre los que destacan los anticuerpos antinucleares (ANAs) dirigidos frente a autoantígenos nucleares como el ADN de doble cadena (ADNdc), las histonas, los nucleosomas y las ribonucleoproteínas. Los ANAs son muy característicos del LES ya que están presentes en cerca del 100% de los pacientes<sup>11</sup>. Sin embargo, no son específicos de la enfermedad y hasta un tercio de la población general puede presentarlos<sup>4, 12</sup>. Algunos autoanticuerpos pueden estar presentes antes de que se diagnostique la enfermedad, y otros se asocian con fenotipos clínicos<sup>13-15</sup>. Por ejemplo, los autoanticuerpos frente a ADN de doble cadena (anti-dsDNA) se asociaron con enfermedad renal y los

autoanticuerpos anti-Ro/SSA y anti-La/SSB están asociados con manifestaciones dermatológicas y lupus neonatal<sup>16</sup>. También se ha encontrado que los autoanticuerpos anti-dsDNA y frente proteínas de unión al ARN (anti-RBP) están asociados con niveles altos de actividad interferón  $\alpha$  (IFN $\alpha$ )<sup>17, 18</sup>. Por otro lado, los autoanticuerpos frente a fosfoproteínas ribosomales (anti-RP) se han asociado a manifestaciones neuro-psiquiátricas (psicosis, ansiedad, disfunción cognitiva y delirio) además de con nefritis, hepatitis y manifestaciones dermatológicas<sup>19</sup>.

**Tabla 1: Criterios de clasificación del LES establecidos por la ACR en 1997<sup>9</sup>.**

<b>Criterio</b>	<b>Definición</b>
<b>1. Eritema malar.</b>	Eritema fijo, plano o elevado, sobre las eminencias malares respetando los pliegues nasolabiales.
<b>2. Eritema discoide.</b>	Zonas eritematosas elevadas con escamas queratóticas adherentes y con taponamiento folicular. En las lesiones antiguas puede producirse cicatrización atrófica.
<b>3. Fotosensibilidad.</b>	Erupción cutánea desproporcionada tras la exposición a la luz solar.
<b>4. Úlceras orales.</b>	Úlceras orales o nasofaríngeas normalmente indoloras.
<b>5. Artritis.</b>	Artritis no erosiva en dos o más articulaciones periféricas, con inflamación, derrame sinovial o dolor a la palpación.
<b>6. Serositis.</b>	Pleuritis o pericarditis.
<b>7. Nefropatía.</b>	<b>Proteinuria</b> persistente superior a 0,5 g/día. <b>Cilindruria:</b> de hematíes o hemoglobina, cilindros granulosos, tubulares o mixtos.
<b>8. Alteración neurológica.</b>	Convulsiones o psicosis, en ausencia de trastorno metabólico, electrolítico o de fármacos que las puedan producir.
<b>9. Alteración hematológica</b>	<b>Anemia hemolítica</b> con reticulocitosis o <b>Leucopenia</b> < de 4.000/mm <sup>3</sup> en 2 ocasiones o <b>Linfopenia</b> < de 1.500/mm <sup>3</sup> en 2 ocasiones o <b>Trombocitopenia</b> < de 100.000/mm <sup>3</sup> no secundaria a fármacos.
<b>10. Alteración inmunológica</b>	<b>Anti-ADNdc</b> positivo ó <b>Anti-Smith</b> positivo ó <b>Anticuerpos antifosfolípidos</b> positivos basado en: 1) Anticuerpos anticardiolipinas (IgG o IgM). 2) Anticoagulante lúpico. Falso positivo en el test serológico para la <b>sífilis</b> ( <i>Venereal Disease Research Laboratory</i> )
<b>11. Anticuerpos antinucleares positivos.</b>	Título anormal de anticuerpos antinucleares (ANAs) por inmunofluorescencia o por otro test equivalente en ausencia de fármacos capaces de producir lupus inducido por los mismos.

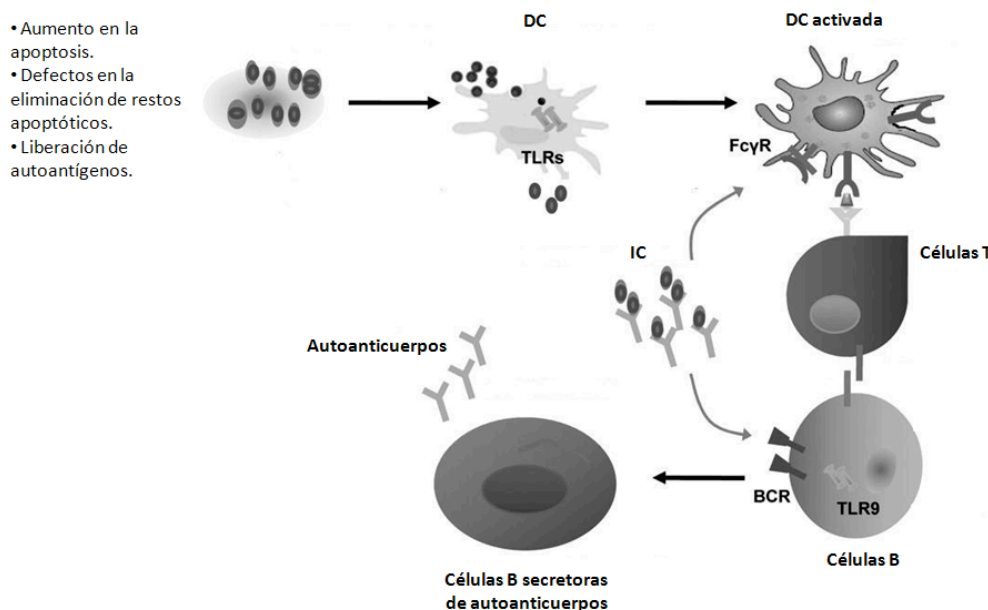
Otro grupo de autoanticuerpos lo forman los que reconocen moléculas de la membrana plasmática como los autoanticuerpos antifosfolípido (aPL). Los aPL están presentes en un 16-60% de los pacientes con LES y están relacionados con el síndrome antifosfolípido, caracterizado por trombosis recurrente y abortos de repetición<sup>11, 20</sup>.

### **1.2. Patogenia del LES:**

Aún no se conocen con detalle los mecanismos patogénicos del LES. Se han implicado factores de predisposición que alterarían tanto el sistema inmune innato como el adquirido dando lugar a la pérdida de la tolerancia inmunológica. Algunos de los posibles componentes patogénicos de la enfermedad son:

- Respuestas inflamatorias persistentes.
- Defectos en la eliminación de restos celulares apoptóticos y de inmuno-complejos (ICs).
- Incremento de la carga autoantigénica.
- Defectos en las proteínas del sistema del complemento (C1q, C2 y C4).
- Derregulación de la presentación de antígenos por las células presentadoras de antígenos (CPAs).
- Alteraciones en la función moduladora de los linfocitos T.
- Anomalías en la producción de autoanticuerpos por los linfocitos B.

De forma tentativa, se ha descrito el curso de la enfermedad como una serie de etapas. Un mecanismo temprano sería la elevada tasa de apoptosis de linfocitos, monocitos y macrófagos. A éste fenómeno se le suman la deficiente eliminación de los restos apoptóticos y un reconocimiento anormal de los autoantígenos liberados durante este proceso por las CPAs. Alteraciones en la regulación de la respuesta inmune sumados a los defectos en presentación de autoantígenos, dan lugar a la activación de linfocitos T y B autoreactivos que producen autoanticuerpos. Los linfocitos autoreactivos infiltran los tejidos, al tiempo que se depositan ICs los cuales desencadenan procesos inflamatorios. Estos dos procesos provocan daño tisular y, por último, fibrosis (**Figura 2**).



**Figura 2: Esquema resumen de los mecanismos patogénicos del LES.** DC: células dendríticas. TLRs: receptores *Toll like*. FcγR: receptores de la fracción constante de la IgG. IC: inmuno-complejos. BCR: receptor de células B. TLR9: receptor *Toll like* 9.

### 1.3. Papel del interferón de tipo I (IFN-I) en la patogenia del LES:

El IFN $\alpha$  es un factor central en la patogenia de la enfermedad. Hace más de 20 años que se observaron niveles elevados de IFN-I en el suero de pacientes con LES<sup>21-23</sup>. Concretamente, el nivel de IFN $\alpha$  se ha encontrado especialmente elevado en las fases de actividad de la enfermedad<sup>24</sup>. Estudios recientes, han mostrado que estos niveles altos de IFN $\alpha$  en el suero de los pacientes constituyen una característica determinada genéticamente<sup>25</sup>. Debido a todas estas evidencias, el IFN $\alpha$  está considerado como un nuevo biomarcador del LES<sup>26, 27</sup> y como diana terapéutica que ya está siendo evaluada<sup>28-32</sup>.

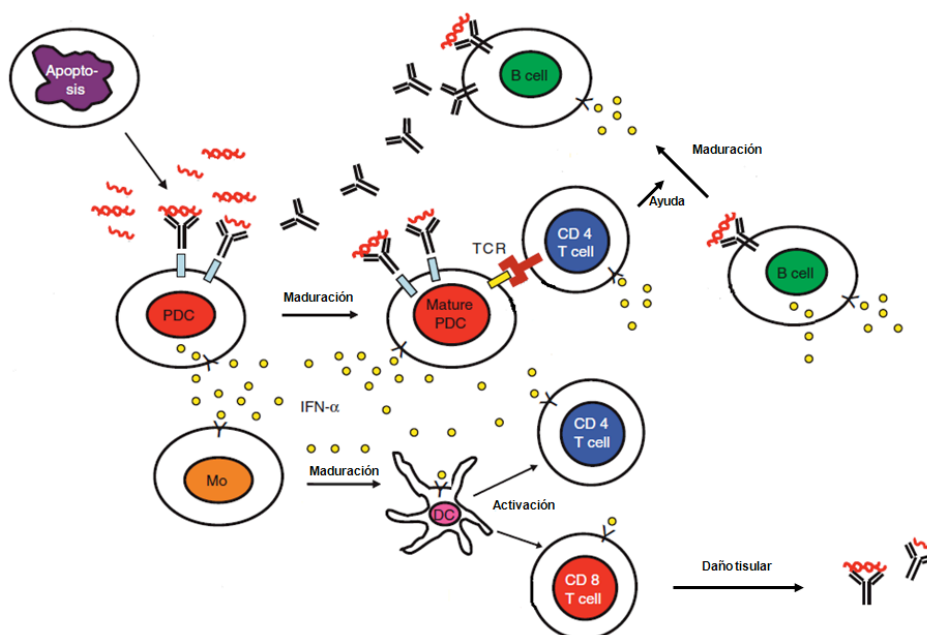
La familia del IFN fue descubierta por su habilidad de interferir en la replicación viral<sup>33</sup>. Las citoquinas del IFN se clasifican en tres familias (de tipo I, II y III) en base a su secuencia de aminoácidos y a los receptores que reconocen. La familia del IFN-I es la más grande y está agrupada en un *cluster* de genes en el cromosoma 9p22. Existen al menos 13 genes que codifican para el IFN $\alpha$ , además de los genes que codifican para el IFN $\beta$ , IFN $\kappa$ , IFN $\tau$  e IFN $\omega$ . Todos los miembros de la familia IFN-I se unen a dos únicos receptores, IFNR1 e IFNR2, e inician la ruta de las quinasas JAK (*Janus* kinasa) /STAT (transductor de señal y activador de la transcripción) estimulando la transcripción de múltiples genes. Los genes inducidos por IFN de tipo I se han agrupado bajo el nombre de la “firma del IFN” que se identificó como característica de los pacientes con LES en estudios con *microarrays*<sup>22</sup>.

Las células dendríticas plasmocitoides (pDCs) son el principal tipo celular productor de IFN-I en el organismo. Estas células producen IFN $\alpha$  en respuesta al reconocimiento de componentes víricos por los receptores *toll-like* (TLRs) TLR7 y TLR9. Estos receptores también reconocen ICs formados por ADN o ARN y autoanticuerpos. Por lo tanto, las pDCs participan tanto en la respuesta inmune innata como en la repuesta inmune adaptativa<sup>34</sup>. Estas células constituyen menos de un 1% de las PBMCs. Sorprendentemente, su frecuencia está disminuida hasta en 100 veces en la sangre de pacientes con LES. Esta disminución se ha interpretado como debida a su masiva infiltración de los tejidos inflamados, como los riñones y la piel<sup>29</sup>. También producen IFN de tipo I los linfocitos B al ser estimulados de forma paralela a través del BCR y de los TLR por el ADN y el ARN liberados.

Como ya se ha mencionado, los TLRs son esenciales en la vía IFN-I. La familia de los TLRs comprende diez miembros que reconocen patrones moleculares característicos de patógenos: ARN de doble cadena (ARNdc), lipopolisacáridos (LPS), proteínas flagelares y algunos lipopéptidos. Los receptores TLR7, el cual reconoce ARNcs, y el receptor TLR9, el cual reconoce secuencias CpG demetiladas de ADN de cadena sencilla, son los principales receptores expresados en pDCs<sup>35,36</sup>. Estos dos TLRs también se expresan en los linfocitos B<sup>37</sup>.

Se han postulado varias causas de los niveles elevados de IFN-I en pacientes con LES. Una de ellas consiste en la acumulación de ICs debido al aumento de la apoptosis y/o defectos en la eliminación de los productos de las células apoptóticas en la sangre de pacientes con LES unido a anomalías en los mecanismos de tolerancia inmunológica. La segunda posibilidad consiste en que una infección vírica puede ser el desencadenante inicial del LES en parte gracias a su potente activación de la vía del IFN-I mediada por las pDCs. Por último, otra posible causa es la existencia de mutaciones presentes en las proteínas de la ruta de señalización del IFN-I que facilitarían su activación<sup>29,30</sup>.

Gracias a la regulación de la expresión de múltiples genes, el IFN-I presenta capacidad immuno-moduladora de la activación de varios tipos de células del sistema inmune<sup>38</sup> (**Figura 3**). Por acción del IFN $\alpha$ , los monocitos se diferencian hacia células dendríticas mieloides (mDCs) presentadoras de autoantígenos. Estas mDCs maduras capturan los restos de las células apoptóticas presentes en la sangre de los pacientes y las presentan a los linfocitos T. A su vez, las mDCs son activadas por el IFN $\alpha$  y presentan los autoantígenos de forma eficiente y acompañada de moléculas coactivadoras en su superficie activando a las células T autoreactivas. Por lo tanto, el IFN $\alpha$  induce y mantiene la generación de mDCs las cuales activan a las células T autoreactivas.



**Figura 3: Esquema resumen del papel del IFN-I en la patogénesis del LES<sup>39</sup>.** PDC: células dendríticas plasmocitoides. Mo: monocitos. DC: células dendríticas mieloideas. TCR: receptor de células T.

Por otro lado, el IFN $\alpha$  junto con la IL6 promueve la diferenciación de los linfocitos B autoreactivos en células secretoras de autoanticuerpos. A su vez, las células B autoreactivas producen IFN $\alpha$  como ya se mencionó. Al mismo tiempo, el IFN $\alpha$  promueve la expresión del activador de células B (BLyS/BAFF) por parte de las pDCs el cual promueve la maduración y supervivencia de las células B maduras productoras de autoanticuerpos. Además, las células T colaboradoras CD4+ autoreactivas (Th) que son activadas por el IFN $\alpha$  y por el reconocimiento mediante el receptor de células T (TCR) de autoantígenos presentados por las mDCs, promueven la maduración de las células B productoras de autoanticuerpos<sup>39</sup>.

El aumento de la apoptosis y la producción de autoanticuerpos promueven la formación de ICs que, a su vez, estimulan la producción de IFN $\alpha$  por las pDCs. Los ICs que activan a las pDCs tras su internalización por el receptor de membrana FCGR2A (receptor IIA del Fc de la IgG)<sup>40</sup>. Después de ser internalizados, los ICs son reconocidos por los receptores TLR7 y TLR9 en los endosomas activando la cascada de señalización de expresión del IFN $\alpha$ <sup>41</sup>. Por lo tanto, se produciría así un efecto de retro-alimentación positiva mediante el aumento nuevamente de la expresión de IFN $\alpha$  por parte de las pDCs.

Una de las preguntas sin resolver es por qué no se agota la producción de IFN $\alpha$  en pacientes con LES. Una posible respuesta es que los altos niveles de IFN $\alpha$  provocan también una

reducción en el número o una activación defectuosa de los linfocitos T reguladores (Treg) y de las células NK<sup>42-46</sup>. Esto permitirá mantener activadas las células T y B autoreactivas.

#### **1.4. Etiología del LES:**

La etiología del LES está lejos de ser bien conocida, pero parece muy compleja y en ella están involucrados factores ambientales, factores genéticos y factores hormonales. La participación de un componente genético está ampliamente demostrada. En primer lugar, el LES presenta agregación familiar, esto es, que los familiares de los pacientes con LES presentan mayor riesgo de desarrollar la enfermedad que el resto de la población. Así, el riesgo relativo de padecer la enfermedad de los hermanos ( $\lambda_s$ ) de un paciente es 20-40 veces mayor que el riesgo en la población general. Por otra parte, la enfermedad es concordante entre gemelos en el 2-5% de los dizigotos y en el 24-56% de los monozigotos lo que se comprende con una heredabilidad importante, de más del 66%<sup>47-53</sup>.

Se interpreta como una evidencia importante de que los factores ambientales están involucrados en la etiología del LES el que la concordancia entre gemelos monozigotos sólo llega a un máximo del 56%. El componente ambiental posiblemente incluye factores tales como agentes mutagénicos (luz ultravioleta), infecciones víricas (virus *Epstein-Bar*) además de la dieta y el estilo de vida (obesidad, hábito de fumar). Estos factores modularían la susceptibilidad al LES alterando los procesos de apoptosis o inmunidad, o a través de cambios epigenéticos muy ligados a cambios en el ambiente. Por último, en relación a los factores hormonales, se ha propuesto que los estrógenos son promotores de la enfermedad en relación con la mayor prevalencia del LES en mujeres (9:1, mujer:hombre)<sup>54-56</sup>.

## **2. Estudio de la genética del LES.**

Se ha encontrado que, con la excepción de algunas familias con herencia mendeliana, en la predisposición genética al LES contribuyen múltiples *loci* de baja penetrancia, por lo tanto, cada factor genético involucrado tiene una aportación débil a la enfermedad. Por lo tanto, el desarrollo de la enfermedad tendría lugar cuando el contenido acumulado de alelos de susceptibilidad excede un umbral, y sería desencadenado por procesos no genéticos de naturaleza poco clara incluyendo algunas infecciones víricas.



En la búsqueda de factores genéticos de predisposición al LES se distinguen dos estrategias de estudio: los estudios de ligamiento de genoma completo y los análisis de asociación; dentro de éstos últimos desatacaremos los estudios de asociación a nivel de genoma completo (GWAS, *genome wide association studies*)

### **2.1. Estudios de ligamiento:**

Los estudios de ligamiento de genoma completo analizan la co-segregación de marcadores genéticos multialélicos (principalmente microsatélites) situados a intervalos de 10-15 Kb a lo largo del genoma. La finalidad de estos estudios es identificar regiones ligadas con el riesgo a padecer una enfermedad que comparten los miembros afectados de familias multi-caso. Al estudiar marcadores repartidos por todo el genoma, si uno de ellos está en ligamiento con un *locus* de susceptibilidad, los miembros enfermos de la familia tendrán el mismo alelo del marcador. De esta manera, se observan las regiones cromosómicas que segregan con la enfermedad y se buscan posibles genes candidatos en las mismas.

Los estudios de ligamiento son el método ideal para el estudio de los fenotipos de herencia mendeliana, ya que los modelos genéticos son pocos y de fácil análisis. También pueden ser una buena aproximación para encontrar algunas variantes raras con grandes efectos, es decir, que provocan una alta probabilidad de desarrollar la enfermedad. Sin embargo, su aplicación a los fenotipos complejos como el LES es más problemática. Esto es debido a que no existe un modelo que explique adecuadamente el patrón de herencia, además de que los factores genéticos de susceptibilidad para estas enfermedades suelen ser muchos y tener un efecto pequeño.

Estudios de ligamiento realizados en familias multi-caso, han permitido identificar numerosas regiones cromosómicas o *loci* de susceptibilidad a la enfermedad con uno o varios genes de predisposición en cada región. Se han publicado 12 estudios de ligamiento de genoma completo con colecciones de familias multi-caso caucásicas, hispanas, afro-americanas y asiáticas<sup>57-68</sup>. En conjunto, casi 60 *loci* han mostrados evidencias de ligamiento con LES, pero solamente 17 han sido replicados en, al menos, 2 estudios independientes: *1p36*, *1q23-24*, *2q23-33*, *2q37*, *4p15-13*, *4q32*, *5p15*, *6p22-q15*, *11q23*, *12q24*, *14q23*, *16p13*, *16q12-13*, *17q11*, *18q21*, *19q13.1* y *20p13*. Con el fin de resumir estos resultados, se han realizado 2 meta-análisis (GSMA, *genome-search meta-analysis*) incluyendo varios estudios de ligamiento de genoma completo<sup>69, 70</sup>. Ambos meta-análisis mantienen los *loci* presentes en los cromosomas 6 y 16 como los más significativos. La región *6p11-21* vendría explicada por genes localizados dentro

HLA, mientras que el gen *NOD2* (*nucleotide-binding oligodimerization domain 2*) se propuso como el causante del ligamiento de la región *16q12*. Sin embargo, estudios directos de este gen *NOD2* mostraron que no está asociado con la susceptibilidad a LES.

La mayoría de los *loci* de susceptibilidad detectados en estos estudios no han sido reproducidos consistentemente. Esto puede deberse a varias limitaciones entre las que destacan: el efecto débil de los *loci* detectados, la heterogeneidad genética o fenotípica de las colecciones, el pequeño tamaño muestral de los estudios y el escaso poder de detección del ligamiento. Además, la región que identifican como ligada a la enfermedad es muy grande, lo cual hace muy difícil la detección del polimorfismo responsable de dicho ligamiento. Por todo ello, los estudios de ligamiento han sido relegados a un segundo plano por los estudios de asociación. Éstos últimos tienen un mayor poder estadístico para detectar variantes de baja penetrancia. Además, facilitan la obtención de tamaños muestrales más grandes al no basarse en familias multi-caso<sup>51</sup>.

## **2.2. Estudios de asociación:**

Los estudios de asociación incluyen tanto diseños caso-control, como análisis de desequilibrio de transmisión en familias (TDT, *family-based transmission disequilibrium test*).

Mediante el test TDT se analiza si existe una transmisión preferente de un alelo o haplotipo específico desde los padres heterocigotos a los hijos enfermos<sup>51</sup>. Para que las familias sean informativas, se requieren como mínimo tríos constituidos por el hijo afecto y sus progenitores de los cuales al menos uno de ellos debe ser heterocigoto. Esta característica puede ser una limitación para la obtención de un tamaño muestral suficiente. El análisis estadístico más sencillo de estos estudios consiste en la  $\chi^2$  de *McNemar's* que evalúa la hipótesis nula de herencia mendeliana (los alelos son transmitidos al 50 % de los hijos desde los padres heterocigotos) frente a la hipótesis alternativa (un alelo de susceptibilidad se transmite más frecuentemente a la descendencia afectada).

El otro tipo de aproximación de los estudios de asociación son los diseños caso-control. Estos estudios se basan en la comparación de individuos enfermos no relacionados con individuos sanos de la misma población. Lo cual permite obtener con facilidad grandes tamaños muestrales aumentando así su potencia estadística. El método estadístico más sencillo de estos estudios consiste en el análisis  $\chi^2$  de tablas de contingencia 2 x 2, construidas con las frecuencias

de los dos alelos de un polimorfismo en casos y controles. Por lo tanto, se dice que uno de los alelos está asociado con la enfermedad si éste tiene una frecuencia significativamente mayor en los casos en comparación con los controles, o viceversa. Sin embargo, una de las limitaciones de los estudios de asociación es la elección del grupo control adecuado. En ocasiones, esta elección se complica por estratificación poblacional que consiste en diferencias en la frecuencia de la enfermedad o del polimorfismo entre las subpoblaciones. Este fenómeno puede dar lugar a resultados artefactuales. Sin embargo, los estudios de asociación se han impuesto por su mayor potencia estadística para detectar polimorfismos con efecto pequeño, y por la relativa facilidad de utilizar mayores tamaños muestrales y, por lo tanto, estudios con mayor poder estadístico.

Los dos tipos de estudios de asociación se han utilizado para evaluar el papel de múltiples genes en la predisposición al LES. Por una parte, genes que están involucrados en procesos que participan en la enfermedad (genes candidatos), por otra parte, en el mapeo fino de *loci* ligados a LES. Así, esta metodología permitió reducir el intervalo de ligamiento del cromosoma 6 al segmento cromosómico que contiene los genes HLA de clase II<sup>71</sup>. Por otro lado, otro resultado obtenido con los estudios de asociación, es la posible explicación del *locus* *1q23* por la presencia de dos polimorfismos no sinónimos (el cambio de base produce un codón que codifica otro aminoácido) en los FCGRs: *FCGR2A* (H131R) y *FCGR3A* (F176V). Estas dos modificaciones alteran su afinidad por diferentes subclases de IgG interviniendo así en el procesamiento de los ICs<sup>64</sup>. Un candidato para el ligamiento del *locus* *1q41-42* es la poli-ADP-ribosa polimerasa (*PARP*), que media la ADP-ribosilación post-traducciona de proteínas<sup>72</sup>. Sin embargo, no se ha encontrado asociación consistente con el LES. El ligamiento de la región *2q23-33* podría venir explicado por los genes *STAT1-STAT4*, los cuales codifican para factores de transcripción que inducen la expresión de genes activados por IFN y su asociación se ha confirmado en múltiples estudios<sup>73</sup>. Para el *locus* *2q37*, se propuso al gen de muerte celular programada 1 (*PDCD1*)<sup>74</sup>. Pero no ha sido confirmado en estudios posteriores<sup>75, 76</sup>. Otros genes cuya asociación ha sido establecida mediante este tipo de estudios son *PTPN22*<sup>77-80</sup>, *IRF5*<sup>22, 23</sup> y *TYK2* (*Janus* tirosín quinasa 2)<sup>81</sup>. En un estudio más reciente, se descubrió la asociación con LES del gen de la integrina  $\alpha$ -M (*ITGAM*) en el intervalo de ligamiento *16p12-q13*, que participa en la inflamación al regular la adhesión y migración leucocitaria<sup>82</sup>.

### **2.3. Estudios de asociación de genoma completo:**

Gracias a los rápidos avances de la tecnología molecular, se han publicado en los últimos años estudios de asociación de genoma completo (GWAS). El enfoque de estos estudios

consiste en evaluar la variabilidad genética a lo largo de todo el genoma sin hipótesis previa. Este enfoque vino facilitado por el descubrimiento de que el genoma humano está mayoritariamente dividido en bloques de desequilibrio de ligamiento (LD)<sup>83</sup>. El LD describe una situación en la cual determinadas combinaciones de alelos aparecen con mayor frecuencia en la población de la que se esperaría al azar (a partir de frecuencias alélicas). Por lo tanto, el genoma humano está estructurado en bloques con alto LD, es decir, en regiones con variantes fuertemente correlacionadas. Éste hallazgo permitió explorar la variabilidad del genoma utilizando un número relativamente pequeño de marcadores. Además, la evolución de la tecnología hacia el genotipado a gran escala dio lugar al desarrollo de *chips* para analizar hasta 2,5 millones de SNPs (polimorfismos de una única base), siendo los más utilizados hasta el momento los de 300000 y 500000 SNPs. Las dos compañías que han proporcionado dichas herramientas son *Affymetrix* e *Illumina*.

Las principales ventajas de los GWAS son su potencia estadística, su rigor y su agnosticismo. Esta última característica hace que se puedan considerar como estudios generadores de nuevas hipótesis, ya que se analiza todo el genoma sin hipótesis previas. Gracias a esto, se han descubierto genes de susceptibilidad que no se esperaba que tuviesen un papel en la enfermedad. Su rigor está relacionado con la necesidad de corregir por el número alto de hipótesis independientes analizadas. La solución propuesta para este problema de múltiples tests supuso establecer un umbral de significación mucho más estricto,  $P < 5 \times 10^{-7}$ , que el clásico de  $P < 0.05$ <sup>84</sup>. Bien, considerando un millón de regiones independientes en el genoma, o, el más aceptado,  $P < 5 \times 10^{-8}$ , que corregiría para entre uno a dos millones de tests independientes<sup>85, 86</sup>. Estos dos umbrales reducen la tasa de falsos positivos, pero también hacen que se necesite un tamaño muestral grande para detectar incluso las asociaciones más fuertes. El aumento de rigor, ha convertido en un requisito que todas las señales asociadas sean confirmadas posteriormente mediante estudios de replicación. Por otro lado, la potencia estadística de los GWAS, es decir, la probabilidad de que el GWAS encuentre una asociación significativa depende de los SNPs incluidos en el *chip*, del tamaño del efecto a detectar y del tamaño muestral disponible. En general, los GWAS tienen una buena potencia ( $> 80 \%$ ) para detectar efectos fuertes (*Odds Ratio* (OR)  $> 1.5$ ), pero ésta se ve muy reducida ( $< 70 \%$ ) en la identificación de efectos más modestos (OR  $> 1.2$ ).

Además, los estudios de GWAS han introducido mecanismos necesarios para reducir el efecto de la estratificación poblacional que es el talón de Aquiles de los estudios de asociación caso-control<sup>87</sup>. Se utilizan dos aproximaciones: el control genómico y el análisis de componentes principales. El control genómico consiste en corregir todos los resultados de asociación en el estudio por un factor de inflación denominado  $\lambda$ . Este factor de inflación  $\lambda$  es

una medida obtenida a partir del test  $\chi^2$  considerando la hipótesis nula ( $\lambda=1$ ) como la ausencia de estratificación poblacional. Si la distribución de genotipos de algún SNP no se ajusta a la hipótesis nula, su factor de inflación  $\lambda$  tendrá un valor mayor que 1, indicando un incremento del estadístico de asociación debido a estratificación poblacional<sup>88</sup>. El problema de este método reside en el hecho de que algunos SNPs difieren más que otros entre las distintas poblaciones, entonces el ajuste uniforme de todos ellos con el mismo valor de  $\lambda$  no es óptimo. El segundo método de corrección se basa en las componentes principales. Las componentes principales son combinaciones lineales de la variabilidad de frecuencia de cada SNP normalizada y combinada a través de todo el GWAS. Por lo tanto, este análisis reduce todos los genotipos a un menor número de variables continuas no correlacionadas. Por lo tanto, si existen subpoblaciones en el estudio, es necesario corregir las distribuciones de los genotipos de cada SNP con estas componentes principales. De esta manera se tiene en cuenta la variabilidad atribuida a la ascendencia y los resultados de asociación no presentan interferencia de subestructura poblacional<sup>89,90</sup>.

Para la obtención de unos datos robustos, es necesario introducir controles de calidad a diferentes niveles: muestras, SNPs y genotipos. En cuanto a las muestras, hay que seleccionar tanto a los casos como a los controles. La selección de casos debe minimizar la heterogeneidad de fenotipos clínicos. Esta tarea es especialmente complicada en enfermedades complejas como es el LES, pues múltiples fenotipos se comprenden con una única clasificación. Con respecto a la selección de controles, éstos deben ser representativos de la misma población de los pacientes. En cuanto a las muestras de ADN, éstas deben tener una calidad óptima y similar entre casos y controles ya que los diferentes métodos de extracción pueden introducir sesgos en los resultados. A partir de la obtención de los datos brutos del genotipado, también se comprueba la concordancia entre el sexo recogido y el observado y si hay muestras que puedan tener una relación de parentesco oculta. Ambos artefactos pueden dar lugar a falsos positivos. Respecto a la calidad del genotipado de cada SNP, lo primero que se comprueba es la eficiencia de genotipado de cada uno de ellos. Estas eficiencias deben ser muy altas y similares entre casos y controles para evitar artefactos. Por último, se comprueba que las frecuencias de los genotipos se encuentren en equilibrio de *Hardy-Weinberg*, así como su concordancia con los datos de HapMap y entre los duplicados internos.

Una limitación importante de los GWAS es el SNP asociado no tiene porque ser la variante causal que influencia la enfermedad. Para identificarlas hacen falta estudios de mapeado fino del gen de interés y re-secuenciado ayudados por bases de datos de modificaciones epigenéticas como el proyecto ENCODE (*ENCyclopedia Of DNA Elements*) que pueden proporcionar evidencias de la posible repercusión funcional responsable de la asociación<sup>91, 92</sup>. Los

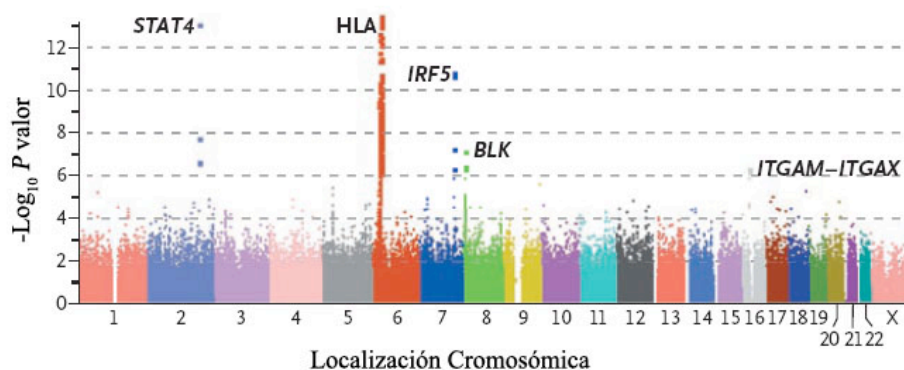
experimentos funcionales posteriores mostrarán cuáles son las variantes causales, aunque en ocasiones este proceso resulta muy laborioso. Además, es posible que un factor de susceptibilidad se escape del umbral de detección de los GWAS. Esto es especialmente posible para las variantes raras (frecuencia del alelo menor  $< 5\%$ ), que no están bien cubiertas por los *chips* utilizados. Por otra parte, los GWAS no pueden evaluar polimorfismos de inserción/delección (indel) o las inversiones. Sin embargo, las últimas plataformas de genotipado ya la detección de los polimorfismos debidos a variaciones en el número de copias (CNVs). Éste avance ha llevado al descubrimiento de su efecto en desórdenes neuropsiquiátricos, T1D, degeneración macular y obesidad<sup>93-96</sup>.

### 2.3.1 GWAS de LES en Europeos:

Hasta la fecha, se han publicado 6 GWAS en LES, cuatro de ellos en población europea<sup>97-100</sup> y dos en asiáticos<sup>101, 102</sup>. Estos estudios no sólo han permitido identificar un gran número de *loci* de susceptibilidad, sino que también han confirmado la asociación con LES de factores genéticos ya conocidos, como HLA, *IRF5*, *FCGR2A*, *PTPN22* y *STAT4*.

El Consorcio Internacional de Genética del LES (SLEGEN) llevó a cabo un estudio de 317501 SNPs en 2 colecciones de muestras con un total de 720 mujeres con LES y 2337 controles. Todos ellos eran norte-americanos con ascendencia europea. Encontraron 8230 SNPs asociados con un nivel sugestivo fueron analizados en otros 2 conjuntos de muestras independientes que consistían en un total de 1846 mujeres con LES y 1825 controles. Los resultados principales ha sido la confirmación de las asociaciones de HLA e *IRF5*, *FCGR2A*, *PTPN22* y *STAT4*. Además, encontraron asociación con 3 SNPs del locus *ITGAM* (rs9888739, rs1143678 y rs4548893), con el SNP rs4963128 del locus *KIAA1542*; con rs6445975 del locus *PXK*, que codifica para una quinasa de serina-treonina con el dominio homóloga *Phox*, y con el SNP rs10798269 en la región *1q25.1* cerca del cual no se encuentra ningún gen conocido. También se encontraron otras asociaciones dignas de mención aunque no cumplían alguno de los criterios de significación o de replicación definidos en este estudio. Entre ellos están: *NMNAT2* (*Nicotinamide mononucleotide adenylyltransferase 2*), *ATG5* (*Autophagy protein 5*), *ICA1* (*Islet cell autoantigen 1*), *XKR6* (*Kell blood group complex subunit-related family, member 6*), *C8orf13-BLK* (*B lymphocyte kinase*), *LYN* (*Tyrosine-protein kinase Lyn*), *UBE2L3* (*Ubiquitin-conjugating enzyme E2 L3*), *SCUBE1* (*signal peptide CUB domain, EGF-like 1*) y la regiones cromosómicas *8p23.1* y *8p21.1*<sup>98</sup> (**Tabla 2**).

Un segundo GWAS analizó 502033 SNPs en 1311 casos con LES y 3340 controles, todos ellos norte-americanos de ascendencia europea. Los resultados más destacables fueron confirmados en 793 casos con LES y 857 controles de población sueca. De igual manera que el estudio anterior, este GWAS confirmó la asociación de *STAT4*, *IRF5* y la región HLA. Sus resultados también mostraron la asociación significativa de polimorfismos en la región cromosómica que contiene el genes *ITGAM-ITGAX* (*Integrin alpha M- Integrin alpha X*) y de la región promotora de los genes *BLK* y *C8orf13* que se transcriben en dirección opuesta<sup>99</sup> (**Tabla 2**) (**Figura 4**).



**Figura 4: Diagrama de Manhattan del GWAS de Hom G. et al<sup>99</sup>.** Los puntos representan todos los SNPs analizados. El eje X muestra su posición en los cromosomas. El eje Y representa el  $-\log_{10}$  del valor *P* de cada SNP.

Otro de los GWAS publicados en 2008, incluyó el estudio de 311238 SNPs en 431 pacientes y 2155 controles norte-americanos de ascendencia europea. En una segunda parte, analizaron 134 *loci* asociados sin llegar a nivel de genoma completo en 740 tríos independientes. También se confirmaron las asociaciones de la región HLA, y de *IRF5* y *STAT4*. Además, se encontró asociación de los genes *BLK* e *ITGAM* (este último sin llegar a alcanzar el nivel de significación de genoma completo). Sin embargo, el resultados más importante de este estudio, es que se encontró otro *locus* con significación a nivel de genoma completo no descrito previamente en LES: la proteína 3 inducida por TNF $\alpha$  (*TNFAIP3*)<sup>97</sup> (**Tabla 2**). Este *locus* ya se había visto asociado con AR previamente<sup>103, 104</sup>.

El último GWAS realizado en europeos hasta la fecha, consistió en un estudio de menor densidad, 85042 SNPs, en 279 casos y 515 controles de origen sueco. Se encontró asociación de la proteína estructural de células B con repeticiones de anquirina (*BANK1*) que alcanzó el valor de  $P < 5 \times 10^{-8}$  en las colecciones adicionales con procedencia escandinava, argentina, alemana, italiana y española (2003 pacientes y 1968 controles en total)<sup>100</sup> (**Tabla 2**).

En un estudio de replicación a gran escala, se analizaron 2466 SNPs que mostraban una  $P < 0.05$  en el GWAS de Hom G. et al.<sup>99, 105</sup>, 505 SNPs de 23 *loci* previamente identificados como asociados con LES y 42 SNPs implicados en otras enfermedades autoinmunes. El genotipado se llevó a cabo en colecciones de Estados Unidos (1129 casos y 2991 controles) y Suecia (834 casos y 1338 controles) independientes a las utilizadas en el GWAS de referencia<sup>99</sup>. Como resultado principal, se confirmaron las asociaciones previas de la región HLA, *STAT4*, *IRF5*, *ITGAM*, *C8orf13-BLK*, *TNFAIP3*, *PTPN22*, *PHRF1* (llamado anteriormente *KIAA1542*) y *TNFSF4*. Con valores de  $P$  entre  $7,5 \times 10^{-07}$  y  $1,5 \times 10^{-03}$ , se confirmaron más *loci* asociados con la susceptibilidad al LES: la región *IRAK1-MECP2* (*methyl CpG binding protein 2*), *PTTG1*, *UBE2L3*, *ATG5*, *BANK1*, *PXK*, *FCGR2A*, *TYK2*, *ICAI* y *NMNAT2*. Por último, en este estudio se identificaron hasta 26 nuevo *loci* de susceptibilidad al LES. Seis de éstos presentaron una asociación a nivel de genoma completo en el análisis combinado: *TNIP1* (*TNFAIP3-interacting protein 1*), *PRDM1* (*PR domain zinc finger protein 1*, *BLIMP-1*), *JAZF1* (*JAZF zinc finger 1*), *UHRF1BP1* (*UHRF1 binding protein 1*), *IL10* (*Interleukin 10*) y *NCF2* (*Neutrophil cytosol factor 2*).

### 2.3.2 GWAS de LES en Asiáticos:

Se han realizado dos GWAS independientes en sujetos de origen asiático. Un primer GWAS se realizó en un chinos<sup>106</sup>. En este estudio se analizaron 493955 SNPs en 1047 casos y 1205 controles y las asociaciones significativas fueron evaluadas en dos colecciones de muestras adicionales con 3152 casos y 7050 controles. Se confirmó la asociación de 5 *loci* fuertemente consolidados en población europea: *TNFSF4*, *HLA-DRB1*, *STAT4*, *TNFAIP3*, *IRF5* y *BLK*. También se observó asociación con LES en otros *loci* no tan consolidados en población europea pero igualmente asociados: *TNIP1*, *PRDM1-ATG5*, *HIC2* (*hypermethylated in cancer 2*)-*UBE2L3* y *IKZF1* (*Ikaros family zinc finger protein 1*). Por último, se identificaron 7 nuevos *loci* de susceptibilidad al LES: *RASGRP3* (*Ras guanyl-releasing protein 3*), *LRRC18* (*leucine rich repeat containing 18*)-*WDFY4* (*WD repeat- and FYVE domain-containing protein 4*), *ETS1* (*Protein C-ets-1*), *SLC15A4* (*solute carrier family 15*) y las regiones cromosómicas 7q11.23, 11q23.3 y 16p11.2. Es importante destacar que todas las asociaciones mencionadas tienen una significación a nivel de genoma completo (**Tabla 2**).



Tabla 2: Resumen de los hallazgos principales de los GWAS en LES.

		SNP	Locus	Chr	<i>P meta-análisis</i>
GWAS europeos	SLEGEn y col. 2008 (N=6728)	rs3131379	<i>MSH5</i> (HLA)	6p21.33	1.71 x 10 <sup>-52</sup>
		rs9888739	<i>ITGAM</i>	16p11.2	1.61 x 10 <sup>-23</sup>
		rs12537284	<i>IRF5</i>	7q32	3.61 x 10 <sup>-19</sup>
		rs4963128	<i>KIAA1542</i>	11p15.5	3.00 x 10 <sup>-10</sup>
		rs6445975	<i>PXK</i>	3p14.3	7.10 x 10 <sup>-9</sup>
		rs7574865	<i>STAT4</i>	2q32	2.8 x 10 <sup>-9</sup>
		rs10798269	-	1q25.1	1.11 x 10 <sup>-7</sup>
		rs1801274	<i>FCGR2A</i>	1q23	6.78 x 10 <sup>-7</sup>
		rs2476601	<i>PTPN22</i>	1p13	6.2 x 10 <sup>-6</sup>
	Hom G y col., 2008 (N=6301)	rs2187668	<i>HLA-DQA1</i> (HLA)	6p21	2.71 x 10 <sup>-21</sup>
		rs7574865	<i>STAT4</i>	2q32.3	8.96 x 10 <sup>-14</sup>
		rs10488631	<i>IRF5</i>	7q32	1.65 x 10 <sup>-11</sup>
		rs13277113	<i>C8orf13-BLK</i>	8p21.3	7.57 x 10 <sup>-8</sup>
		rs11574637	<i>ITGAM-ITGAX</i>	16p11.2	5.35 x 10 <sup>-7</sup>
	Graham RR y col., 2008 (N=2586; y 740 trios)	-	<i>HLA</i>	6p21	< 5 x 10 <sup>-8</sup>
		-	<i>IRF5/TNPO3</i>	7q32	< 5 x 10 <sup>-8</sup>
		rs5029939	<i>TNFAIP3</i>	6q23.3	2.89 x 10 <sup>-12</sup>
		rs3821236	<i>STAT4</i>	2q32.3	8.49 x 10 <sup>-11</sup>
		rs2618476	<i>BLK</i>	8p23.1	1.70 x 10 <sup>-8</sup>
		rs11150610	<i>ITGAM</i>	16p11.2	1.72 x 10 <sup>-6</sup>
	Kozyrev SV y col., 2008 (N=3971)	rs17266594	<i>BANK1</i>	4q24	4.74 x 10 <sup>-11</sup>

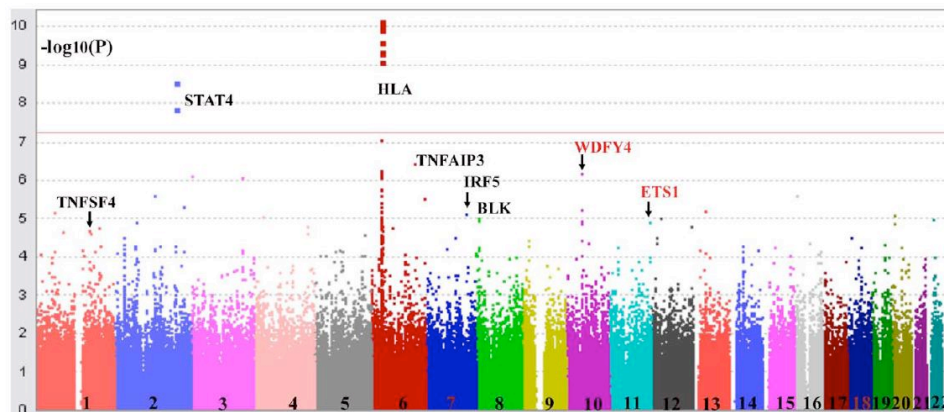
**Tabla 2: Resumen de los hallazgos principales de los GWAS en LES.** Para cada estudio se detallan los SNPs asociados, el *locus* en el que se encuentran, la posición en el cromosoma y el valor de *P* en el análisis conjunto de todos los resultados obtenidos en el mismo estudio.

Continuación tabla 2. Resumen de los hallazgos principales de los GWAS en LES.

GWAS asiáticos		SNP	Locus	Chr	P meta-análisis
Han JW y col., 2009 (N=12454)		rs7574865	STAT4	2q32.3	5.17 x 10 <sup>-42</sup>
		rs2205960	TNFSF4	1q25.1	2.53 x 10 <sup>-32</sup>
		rs6590330	ETS1	11q24.3	1.77 x 10 <sup>-25</sup>
		rs7812879	BLK	8p23.1	2.09 x 10 <sup>-24</sup>
		rs4917014	IKZF1	7p12.2	2.75 x 10 <sup>-23</sup>
		rs4728142	IRF5	7q32.1	8.14 x 10 <sup>-19</sup>
		rs2230926	TNFAIP3	6q23.3	1.37 x 10 <sup>-17</sup>
		rs4639966	Múltiples genes	11q23.3	1.25 x 10 <sup>-16</sup>
		rs463426	HIC2-UBE2L3	22q11.21	1.48 x 10 <sup>-16</sup>
		rs13385731	RASGRP3	2p22.3	1.25 x 10 <sup>-15</sup>
		rs9271100	HLA-DRB1	6p21	1.42 x 10 <sup>-12</sup>
		rs548234	PRDM1-ATG5	6q21	5.18 x 10 <sup>-12</sup>
		rs1913517	LRRC18-WDFY4	10q11.22	7.22 x 10 <sup>-12</sup>
		rs1385374	SLC15A4	12q24.32	1.77 x 10 <sup>-11</sup>
		rs10036748	TNIP1	5q33.1	1.67 x 10 <sup>-9</sup>
		rs1167796	Múltiples genes	7q11.23	2.12 x 10 <sup>-8</sup>
		rs7197475	Múltiples genes	16p11.2	2.77 x 10 <sup>-8</sup>
Yang W y col., 2010 (N=7500)		rs9271366	HLA-DRB/HLA-DQA1	6p21	7.67 x 10 <sup>-10</sup>
		rs7574865	STAT4	2q32.3	4.22 x 10 <sup>-9</sup>
		rs877819	WDFY4	10q11.22	5.66 x 10 <sup>-7</sup>
		rs2736340	BLK	8p23.1	8.56 x 10 <sup>-6</sup>
		rs6590330	ETS1	11q24.3	8.92 x 10 <sup>-6</sup>
		rs729302	IRF5	7q32	1.01 x 10 <sup>-5</sup>
		rs4522865	BANK1	4q24	3.56 x 10 <sup>-4</sup>
		rs704840	TNFSF4	1q25.1	1.21 x 10 <sup>-3</sup>

**Tabla 2: Resumen de los hallazgos principales de los GWAS en LES.** Para cada estudio se detallan los SNPs asociados, el locus en el que se encuentran, la posición en el cromosoma y el valor de P en el análisis conjunto de todos los resultados obtenidos en el mismo estudio.

En el segundo GWAS de población asiática se analizaron 620901 SNPs en un total de 3300 pacientes y 4200 controles procedentes de Hong Kong, China y Tailandia<sup>102</sup>. En este estudio también se replicaron las asociaciones de *loci* de susceptibilidad consolidados en población europea como son: región HLA, *STAT4*, *IRF5*, *BANK1*, *BLK*, *TNFAIP3* y *TNFSF4*. Sin embargo, a diferencia del anterior GWAS en asiáticos, sólo las señales de asociación de la región HLA y de *STAT4* llegaron a una significación a nivel de genoma completo (**Figura 5**). Además, en este GWAS se confirmaron dos *loci* asociados en el GWAS relacionado con chinos<sup>101</sup>: *ETS1* y *WDFY4*.



**Figura 5: Diagrama de Manhattan del GWAS de Yang W. et al<sup>102</sup>.** Los puntos representan todos los SNPs analizados. El eje X muestra su posición en los cromosomas. El eje Y representa el  $-\log_{10}$  del valor  $P$  de cada SNP.

### 2.3.3 La heredabilidad no explicada del LES:

Los GWAS nos han permitido avanzar en la investigación del componente genético del LES ya que se han identificado o confirmado alrededor de 30 factores genéticos de susceptibilidad. Los GWAS se basan en la hipótesis de “enfermedad común, variante común”. Según la cual, el componente genético de las enfermedades comunes se debería a unos cuantos polimorfismos cuyas frecuencias del alelo menor fuesen  $> 5\%$  (variantes comunes). Aunque el LES es una enfermedad rara, los GWAS han tenido un notable éxito en cuanto a la identificación de factores genéticos. Sin embargo, cada uno de los factores genéticos tiene un efecto pequeño y en conjunto sólo exploran una pequeña parte de la heredabilidad. Fenómeno que comparte el LES con otras enfermedades mucho más frecuentes. Así, el estudio de replicación a gran escala<sup>105</sup> que confirmó 26 *loci* de asociación para LES, estimó que sólo explican un 8% de la heredabilidad del LES.

Se han propuesto muchas hipótesis sobre las causas de la heredabilidad no explicada en las enfermedades complejas. Una de ellas es que aún queden muchas variantes comunes de pequeño efecto sin identificar. O que se deba a la heterogeneidad de la enfermedad. Otra posibilidad es que la heredabilidad explicada esté infra-estimada al usar la asociación de los SNPs asociados pero no la de las variantes causales que, muy probablemente, es mayor. Además, es posible que parte de la heredabilidad no explicada se deban a variantes genéticas que no se capturan en los GWAS. Un grupo son las variantes de baja frecuencia (frecuencia alelo menor: 1%-5 %) o variantes raras (frecuencia alelo menor < 1 %). Es posible que algunas de estas variantes poco frecuentes tengan un tamaño de efecto mayor explicando así más proporción de la heredabilidad. Las variantes de baja frecuencia con efecto suficientemente fuerte deberían haber sido identificadas por los estudios de ligamiento, es posible que los estudios no fueran suficientemente potentes o que las variantes tengan un efecto intermedio. Así por ejemplo, hay variantes raras asociadas con fuerte penetrancia al LES y que no aparecieron en los estudio de ligamiento como son las mutaciones en *C1q* y *C2* o en *TREX1* que se comentarán con más detalle en próximos apartados.

Otro tipo son las variantes estructurales, como las CNVs y las inversiones o translocaciones. Estas variantes no se evaluaban de forma explícita en los GWAS. No está claro hasta qué punto este tipo de variantes están cubiertas por los SNPs a través de LD y tampoco cuál es la probabilidad de que afecten a la susceptibilidad al LES. A partir de estudios de genes candidatos se han detectado asociaciones del LES con dos CNVs. Una en el factor del complemento *C4*<sup>107</sup> y otra en *FCGR3B*<sup>108</sup>. Por último, otros posibles contribuyentes a la heredabilidad no explicada son las interacciones gen-ambiente y gen-gen (epístasis)<sup>109</sup>.

### 3. Factores genéticos del LES.

#### 3.1. Región HLA:

El primer factor genético para el LES identificado está en la región del antígeno leucocitario humano (*HLA*) en el cromosoma *6p21.3*. Esta región codifica más de 200 genes muchos de los cuales tienen funciones inmunológicas conocidas. Esta riqueza en genes junto con un alto LD dificulta la identificación de las contribuciones independientes de cada uno de los genes incluidos en esta región. La región HLA está subdividida en tres clases (I, II y III). La regiones HLA de clase I y II contienen genes que codifican para glicoproteínas que procesan y presentan los péptidos para ser reconocidos por las células T. Se conocen unos 20 genes HLA de clase I dentro de los cuales lo más importantes son: *HLA-A*, *HLA-B* y *HLA-C*. Por otro lado, los genes HLA de clase II codifican para las cadenas polipeptídicas  $\alpha$  y  $\beta$  de las moléculas de clase II.

Dentro de esta clase II existen cinco familias: HLA-DM, HLA-DO, HLA-DP, HLA-DQ y HLA-DR. Finalmente, la región HLA de clase III contiene otros genes con función inmunológica como son: *TNF $\alpha$*  (factor de necrosis tumoral  $\alpha$ ), *C2* (componente del complemento 2), *C4A* (componente del complemento 4A), *C4B* (componente del complemento 4B), *CFB* (factor de complemento B), *MSH5* (mutS homólogo 5), *SKIV2L* (helicasa SKIV2L), *CREBL1* (proteína de unión 1 al elemento de respuesta CAMP), *MICB* (polipéptido relacionado con la secuencia B, MHC clase I) and *OR2H2* (receptor olfatorio 2H2)<sup>110, 111</sup>. Como se puede observar, la gran mayoría de las moléculas del HLA están implicadas en el reconocimiento y la respuesta inmune, de ahí que muchos estudios genéticos de enfermedades autoinmunes se centren en ellos.

Uno de los estudios que más contribuyó a definir la asociación del *HLA* con LES incluyó unos 100 microsatélites analizados en 334 familias multicaso. En este análisis se definieron 3 haplotipos asociados con la susceptibilidad al LES. Estos haplotipos incluyen los genes HLA de clase II, *HLA-DRB1* y *DQB1*. Concretamente los haplotipos correspondientes al *HLA-DR2* (*DRB1\*1501-DQB1\*0602*), *HLA-DR3* (*DRB1\*0301-DQB1\*0201*), y con una menor importancia, al *HLA-DR8* (*DRB1\*0801-DQB1\*0402*)<sup>71</sup>. Los haplotipos portadores del *HLA-DR3* confieren mayor riesgo que los portadores del *HLA-DR2* y los individuos homocigotos para DR3 o heterocigotos DR3/DR2 muestran el riesgo más alto. El haplotipo *HLA-DR2* se extiende por unas 500 Kb e incluye toda la región HLA de clase II. El haplotipo *HLA-DR3* se extiende aproximadamente por 1Mb que incluye las regiones HLA de clase II y III. Por último, el *HLA-DR8*, de menor frecuencia en Europa, se extiende unas 500 Kb en la región HLA de clase II. Estos haplotipos se han asociado con la producción de autoanticuerpos característicos del LES<sup>112, 113</sup>. Concretamente, el haplotipo *HLA-DR2* se asocia con la producción de anti-Sm y anti-ADNdc, mientras que *HLA-DR3* está relacionado con anti-Ro y anti-La.

Los GWAS han confirmado que algunos de los SNPs localizados en la región del HLA de clase II están entre los más fuertemente asociados con la enfermedad<sup>97-99, 101, 102, 105, 114</sup>. Sin embargo, su contribución al LES no es mucho mayor que la de otros factores genéticos, por lo que el efecto del HLA no es tan destacado como el observado en otras enfermedades autoinmunes como AR o T1D<sup>115</sup>. El GWAS de SLEGEN<sup>98</sup>, encontró la asociación más importante en europeos en un SNP (rs3131379) localizado en el gen *MSH5* perteneciente a la región HLA de clase III. Este gen está involucrado en las reorganizaciones del DNA que permiten dar lugar a la amplia variedad de inmunoglobulinas. Recientemente, se ha replicado esta asociación en una población de ascendencia afro-americana<sup>116</sup>. Otro de los genes pertenecientes a la región HLA de clase III, con un posible efecto independiente en la susceptibilidad al LES, es *SKIV2L* identificado en un estudio de asociación basado en familias del Reino Unido<sup>115</sup>. Este gen codifica para una helicasa de ARN y se expresa en células T, B y

dendríticas. Más recientemente, se ha llevado a cabo un *screening* exhaustivo de toda la región HLA analizando 1974 SNPs en 1610 casos y 1470 controles de ascendencia europea. Como resultado, se ha confirmado la asociación de los genes *HLA-DR* y *HLA-DQ*. Sin embargo, también se han encontrado múltiples posibles asociaciones independientes en genes de la región del HLA de clase III como son: *CREBL1*, *MICB* y *OR2H2*<sup>117</sup>.

Funcionalmente, los genes *HLA-DR* y *HLA-DQ* tienen un papel muy importante en los procesos inmunes ya que su función es presentar péptidos a los receptores de células T (TCR) de las células T colaboradoras T (Th) CD4+. Estas células estimuladas inducen la producción de anticuerpos por las células B (**Figura 6**). Sin embargo, no se conoce porqué unos alelos HLA aumentan la susceptibilidad al LES.

### **3.2. Factores del Complemento:**

Las deficiencias en los factores del complemento *C2*, *C4* y *C1q* muestran una asociación muy fuerte con LES pero no tienen una gran repercusión a nivel poblacional pues son raras o muy raras<sup>118-121</sup>. Los genes *C2* y *C4* están localizados en la región del HLA de clase III, mientras que el gen *C1q* está localizado en la región cromosómica entre *1p34.1-1p36.3*. Una de las funciones del complemento en la que estos factores son necesarios, consiste en la eliminación de restos celulares apoptóticos como de necrosis y de ICs mediante opsonización. Este es un proceso importante en la patogenia del LES ya que los restos celulares pueden funcionar como autoantígenos y los ICs activan procesos inflamatorios y provocan daño tisular<sup>122</sup> (**Figura 6**). Otras funciones incluyen potenciar la respuesta inflamatoria, facilitar la fagocitosis y dirigir la lisis de células infectadas.

La deficiencia completa de *C4* es rara pero se asocia fuertemente con LES<sup>123</sup>. Entre el 54 y el 75% de los individuos desarrollan LES<sup>124</sup>. Sin embargo, la deficiencia parcial es más común debido a un CNV que afecta a los dos genes parálogos que codifican para *C4* y que están situados en la misma región del HLA. Un número bajo de copias se ha asociado con un aumento del riesgo de LES y un número elevado con una disminución del riesgo<sup>125, 126</sup>. Sin embargo, la deficiencia parcial se encuentra en el mismo haplotipo del HLA-DR3, también asociado a LES, por lo que no está claro si la asociación con LES de la deficiencia parcial es o no independiente de la del HLA-DR<sup>122</sup>.

Por otra parte, la deficiencia del gen *C2*, que es la más frecuente en la población general (0,01%), se asocia con LES en el 10-33 % de los individuos<sup>122, 124</sup>. Finalmente, más del 90 % de los sujetos deficientes en *CIq* desarrollan LES, normalmente con un fenotipo severo aunque se trata de una deficiencia rarísima<sup>122</sup>. Por otra parte, varios SNPs de *CIq* se han asociado con nefritis y fotosensibilidad en poblaciones hispanas y afro-americanas<sup>127</sup>, o con manifestaciones cutáneas en europeos<sup>128</sup>.

### 3.3. *TREX1*:

Otro gen en el que se han identificado variantes raras asociadas con LES es *TREX1* que codifica para una 3'-5' exonucleasa del ADN. *TREX1* degrada rápidamente el ADN dañado en las células apoptóticas previniendo la activación inapropiada del sistema inmune<sup>129, 130</sup>. Además, *TREX1* también degrada ADN de cadena sencilla derivado de virus y de retroelementos endógenos reduciendo la repuesta del IFN-I<sup>130</sup> (**Figura 6**).

Inicialmente, las mutaciones en *TREX1* fueron identificadas en pacientes con el síndrome de Aircadi-Goutieres<sup>131</sup>. Este síndrome consiste en una encefalopatía caracterizada por niveles altos de IFN $\alpha$  en el fluido cerebroespinal, *rash* cutáneo, artritis, úlceras orales y ANAs que es similar a una forma poco frecuente de LES, denominada lupus sabañón. Con este antecedente se secuenció *TREX1* en 417 pacientes con LES e identificándose tres variantes raras en 9 que estuvieron ausentes en los 1712 controles<sup>132</sup>.

En un estudio más reciente se han genotipado todas las variantes raras descritas en 8732 pacientes con LES y 7492 controles de múltiples grupos étnicos<sup>133</sup>. Las variantes raras mostraron una frecuencia acumulada del 0.5% en pacientes con LES y mucho más raras en controles. Por otro lado, estos autores encontraron algunas asociaciones entre polimorfismos comunes de este gen y subfenotipos del LES.

### 3.4. *FCGRs*:

Los *FCGRs*, consisten en la familia de genes que codifican para los receptores de baja afinidad de la fracción constante de la inmunoglobulina G (IgG). Están involucrados en repuestas mediadas por anticuerpos como es el reconocimiento de los ICs. Los estudios de ligamiento

proporcionaron las primeras evidencias de la relación con LES del *locus* del cromosoma 1 donde se encuentran los genes que codifican para los *FCGRs*<sup>57</sup>. La asociación más clara corresponde al polimorfismo no sinónimo dentro del gen *FCGR2A* (rs1801274, H131R). El alelo mutante asociado al LES da lugar a la arginina, codificando un receptor con baja afinidad por las partículas opsonizadas con la IgG2 y reduciendo así la degradación de los ICs<sup>134</sup>. Se han publicado más de 20 estudios de este SNP con resultados de asociación con LES y/o nefritis lúpica inconsistentes en distintos grupos étnicos incluyendo colecciones europeas, afro-americanas y koreanas<sup>135-139</sup>. Además, un meta-análisis ha mostrado un efecto moderado de susceptibilidad<sup>140</sup>. A pesar de esta controversia en los resultados previos, esta señal se confirmó como asociada al LES en uno de los GWAS realizado en población europea<sup>98</sup>. A partir de ahí, estos resultados fueron ampliamente confirmados en un estudio de replicación a gran escala y un posterior meta-análisis<sup>105, 114</sup>. Además, otro estudio reciente en población afro-americana también replicó esta asociación<sup>116</sup>. En contraposición a estos resultados, esta asociación no se observó en ninguno de los dos GWAS realizados hasta la fecha en poblaciones asiáticas<sup>101, 102</sup>.

Por otro lado, el polimorfismo no sinónimo de *FCGR3A* (rs396991, F176V) ha mostrado igualmente asociación al LES en algunos estudios y también modifica la afinidad del receptor por los ICs que contienen IgG1, IgG3 o IgG4. El alelo mutante asociado al LES da lugar a la fenilalanina, codificando un receptor con baja afinidad por los ICs reduciendo así su degradación<sup>141, 142</sup>. Un meta-análisis mostró que está asociado pero sólo en los pacientes con nefritis lúpica<sup>140, 143</sup>. Estos dos genes, *FCGR2A* y *FCGR3A*, están separados por tan solo 35 kb, pero las dos asociaciones parecen independientes. Además, las IgG2 y IgG3 son las subclases más ampliamente encontradas en las biopsias de riñón de pacientes con nefritis lúpica<sup>144</sup>. Por lo tanto, la presencia de múltiples alelos de riesgo en estos dos genes, disminuiría de manera acumulativa la degradación de los ICs dando lugar a una interacción genética que aumentaría el riesgo a padecer LES<sup>145</sup>.

También se conoce la existencia de un SNP no sinónimo en el gen *FCGR2B* (rs10917661; I232T) el cual se ha encontrado asociado al LES en población asiática<sup>146-148</sup> pero no en poblaciones ascendencia europea<sup>149-151</sup>. El alelo mutante que da lugar a una treonina reduce la capacidad de inhibición de activación de los linfocitos B por parte de este receptor transmembrana, promoviendo una respuesta autoinmune<sup>152</sup>. Además, se ha encontrado que en población con ascendencia europea existe un haplotipo asociado al LES definido por los alelos del promotor -386G/-120T que produce un aumento de la transcripción del gen *FCGR2B*<sup>148</sup>.

Finalmente, el gen *FCGR3B* es otro gen de esta misma familia de receptores que está implicado en la susceptibilidad al LES. Este gen presenta un polimorfismo de tipo CNV,



observándose que el bajo número de copias está asociado con riesgo de padecer LES ya que promueve la disminución de la expresión del receptor lo que conlleva deficiencias en la adherencia e incorporación de los ICs por los neutrófilos<sup>153, 154</sup>.

A modo de resumen, podemos decir que la función de los FCGRs consiste en unirse a los ICs, por la región Fc del anticuerpo, de modo que las células que expresan estos receptores pueden identificarlos para su eliminación mediante fagocitosis. Concretamente, *FCGR3A* se expresa en células NK, macrófagos y algunos monocitos. Mientras que *FCGR2A* se expresa más ampliamente en las células hematopoyéticas. A su vez, *FCGR2B* y *FCGR3B* se expresan principalmente en linfocitos B y neutrófilos respectivamente. Como papel en la patogenia del LES, se piensa que la baja afinidad de los alelos de riesgo por los complejos inmunes provoca una deficiencia en su eliminación, aumentando la probabilidad de una respuesta autoinmune persistente que contribuiría al desarrollo de la enfermedad (**Figura 6**). Por tanto, la participación de los *FCGRs* así como del sistema del complemento en la susceptibilidad al LES indica la importancia que tiene la deficiente eliminación de ICs en el desarrollo del LES.

### 3.5. TYK2:

El gen *TYK2* codifica para la tirosín quinasa 2 perteneciente a la familia JAK (quinasas *Janus*) formando parte de la ruta JAK-STAT. Por lo tanto, la quinasa TYK2 está involucrada en la regulación de la señalización de IFN $\alpha$ , IL6, IL10 o IL12. Funcionalmente, la quinasa TYK2 es activada por el receptor del IFN $\alpha$  (IFNAR), tras unión de su ligando en PBMCs. Posteriormente, se produce el reclutamiento y activación de las STATs las cuales dimerizan y se translocan al núcleo estimulando la expresión de genes incluidos dentro de la “firma del interferón” anteriormente comentada<sup>22</sup> (**Figura 6**).

Varios polimorfismos dentro del gen *TYK2* se han asociado con la susceptibilidad al LES, pero todos ellos de una manera ambigua dando lugar a una asociación controvertida. Este *locus* se estudió junto con *IRF5* un estudio en población escandinava, que analizaba genes de la ruta del IFN. El polimorfismo más asociado de este estudio fue nsSNP rs2304256 localizado en el exón 8 del gen. Este nsSNP rs2304256 codifica para un cambio de valina por fenilalanina en el dominio homólogo Janus 4, pero su posible efecto funcional aún no ha sido identificado<sup>155</sup>. Sorprendentemente, ningún polimorfismo dentro del este gen mostró asociación en ninguno de los GWAS realizados para LES tanto en población como ascendencia europea como asiática<sup>97-102, 114</sup>. Sin embargo, existen varios estudios independientes, todos ellos en población europea,

que confirman la asociación de *TYK2* con la susceptibilidad al LES<sup>105, 156-158</sup>. Por el contrario, otros estudios muestran resultados negativos con poblaciones tanto de descendencia europea como asiática<sup>159-162</sup>.

### 3.6. *STAT4*:

El gen *STAT4* codifica para un factor de transcripción de la familia STAT (señales de transducción y activación de la transcripción). La asociación de *STAT4* con la susceptibilidad al LES ya se ha descrito mediante estudios de ligamiento, siendo el gen responsable de la asociación de la región cromosómica 2q. Concretamente, esta región se encontró asociada tanto en LES como en AR<sup>57, 163</sup>. Además, este gen también se encontró asociado, tanto con LES como con AR, en posteriores estudios de asociación caso-control independientes en población euro-americana<sup>73, 164</sup>. En estos estudios se identificó que el alelo menor del SNP rs7574865, localizado en el tercer intrón de *STAT4*, confería un fuerte riesgo al LES con valores de OR de entre 1.5 y 1.7. Una confirmación importante de la asociación del gen *STAT4* con LES consiste en que se ha encontrado asociado en tres de los cuatro GWAS realizados con colecciones con ascendencia europea además de en los dos GWAS con colecciones asiáticas<sup>97-99, 101, 102</sup>. De la misma manera, múltiples estudios posteriores confirman la asociación de este gen *STAT4* con la susceptibilidad al LES en población europea y asiática<sup>158, 165-167</sup>. Además, esta asociación se confirmó ampliamente mediante estudios de replicación y meta-análisis tanto en población europea como asiática<sup>105, 114, 168</sup>. Por último, un reciente estudio de genes candidatos también confirma esta asociación en población afro-americana<sup>116</sup>.

En estudios dirigidos hacia la búsqueda del polimorfismo causal de *STAT4* involucrado en la susceptibilidad al LES, se han encontrado diferentes marcadores con asociaciones independientes al SNP rs7574865 y relacionados con diferencias de expresión del *locus*<sup>169-171</sup>. Concretamente, se ha sugerido la existencia de una asociación independiente en el intrón 16, siendo el polimorfismo más asociado el SNP rs3821236. En este mismo estudio, también se observó que solamente los SNPs rs7574865 y rs3821236 se correlacionan con los niveles de expresión de *STAT4*, pero se desconoce cuál es el mecanismo de esta regulación<sup>169</sup>. Recientemente, la asociación de este SNP rs3821236 se ha replicado en una colección canadiense<sup>159</sup>. Por otro lado, también se identificó una nueva señal independiente localizada en el intrón 3 en un estudio de casi 10000 pacientes con LES procedentes de múltiples grupos étnicos (afro-americanos, hispano-americanos, asiáticos y europeo-americanos). Esta señal está

representada por un haplotipo de riesgo que se extiende unas 73 Kb desde el intrón 3 hasta el exón 17 del gen *STAT4*<sup>159, 170</sup>.

Otro hallazgo que le aporta solidez a la asociación del gen *STAT4* con la susceptibilidad al LES, consiste en la asociación de algunas de las manifestaciones clínicas de esta enfermedad con los alelos de riesgo de *STAT4*. El alelo de riesgo del SNP rs7574865 está asociado con un fenotipo del lupus más severo caracterizado por edad temprana de aparición de la enfermedad (< 30 años), nefritis, anticuerpos anti-ADN de doble cadena, anticuerpos aPL, anticuerpos anti-Ro, desorden hematológico, úlceras orales y por último isquemia cerebro-vascular<sup>164, 165, 167, 170-174</sup>. Sin embargo, ninguna de estas asociaciones fueron concordantes en todos los estudios por lo que el fenotipo asociado está todavía por definir totalmente.

Funcionalmente, el gen *STAT4* codifica para un factor de transcripción expresado en varios tipos celulares (linfocitos, macrófagos y DCs). Dentro de sus funciones podemos encontrar diferenciación celular, proliferación y apoptosis. El factor de transcripción STAT4 se encuentra de forma latente en el citosol, pero tras su activación por citocinas, se fosforila y forma homodímeros que se traslocan al núcleo. Dentro de las citoquinas que lo activan están IFN- I y sobre todo IL12, los cuales inducen su fosforilación. Es importante destacar que el alelo de riesgo de este SNP rs7574865 también se ha encontrado asociado con una mayor sensibilidad a la señalización mediada por IFN $\alpha$  en PBMCs<sup>175</sup>. Siguiendo su ruta de señalización, el factor de transcripción STAT4 promueve la expresión múltiples genes como INF $\gamma$  e IL17 en el núcleo de la célula<sup>176, 177</sup> (**Figura 6**). Una posible explicación del papel funcional de este gen vendría dada por medio de la interacción entre los genes *STAT4* e *IRF5* pertenecientes ambos a la ruta del IFN-I. Partiendo de esta hipótesis, en un estudio previamente nombrado, se observó que seleccionando los pacientes portadores del haplotipo de riesgo formado por los alelos de riesgo del SNP rs7582694 de *STAT4* (que define el haplotipo de riesgo y tiene  $r^2 = 1$  con el rs7574865) y de dos polimorfismos asociados en *IRF5* (indel CGGGG y el SNP rs10488631), la OR se incrementaba 1,82 veces<sup>171</sup>. Este mismo fenómeno de incremento del efecto al combinar estos dos factores genéticos se describió en otro estudio de LES con muestras europeas e hispanas<sup>169</sup>. Sin embargo, los aumentos de riesgo observados en estos estudios no parecen ser resultados de una interacción epistática clara ya que no han detectado desviación de la suma de susceptibilidad que confiere cada alelo por sí mismo.

### 3.7. *PTPN22*:

El gen *PTPN22* codifica para una fosfatasa de tirosina linfoide denominada LYP. La función de LYP consiste en inhibir la cascada de señalización del receptor de antígenos de las células T (TCR) mediante su interacción con la Src tirosín-quinasa C-terminal, CSK. En esta interacción, LYP defosforila a las quinasas Lck y Fyn de la familia Src inhibiendo así la respuesta a la señalización mediada por TCR. Por el contrario, CSK fosforila y activa a las quinasas Lck y Fyn de la familia Src activando así la respuesta mediada por TCR<sup>178</sup>. Por lo tanto, es lógico pensar que estas alteraciones en la señalización del TCR puedan incrementar el riesgo de autoinmunidad (**Figura 6**).

Este gen presenta un SNP no sinónimo rs2476601 (Arg620Trp) localizado en el exón 14 e implicado en múltiples enfermedades autoinmunes: T1D, AR, LES y enfermedad tiroidea autoinmune entre otras<sup>179</sup>. La primera evidencia de asociación de este *loci* con LES fue publicada por Kyogoku C et al. en una población con ancestro europeo<sup>79</sup>. A partir de este estudio, la asociación de *PTPN22* con la susceptibilidad al LES se ha confirmado en uno de los GWAS realizados en LES, además de en múltiples estudios independientes en población europea<sup>98, 105, 114, 159, 180, 181</sup>. Con respecto a su asociación en otras poblaciones, ningún polimorfismo dentro del gen *PTPN22* se ha encontrado asociado con la susceptibilidad al LES en los dos GWAS realizados hasta la fecha en población asiática<sup>101, 102</sup>. Sin embargo, en un meta-análisis reciente, en el cual se incluyen varias poblaciones procedentes del norte de Europa, con ascendencia hispana y afro-americanos, se confirma la asociación del nsSNP rs2476601 con LES<sup>182</sup>.

En relación a la repercusión funcional de este nsSNP rs2476601, concretamente su alelo de riesgo 620W impide la interacción de LYP con la tirosín-quinasa CSK. Por lo tanto, se postuló que el alelo de riesgo implicaría una pérdida de función de LYP lo cual llevaría a una disminución en la regulación negativa de los linfocitos T desencadenando una respuesta exagerada y por consiguiente, autoinmunidad. Sin embargo, experimentalmente se ha observado lo opuesto. El alelo de riesgo implica una ganancia de función de LYP, resultando en una mayor inhibición de la señalización del TCR<sup>183</sup>. Por lo tanto, esta hiposensibilidad linfocitaria produce fallos en la eliminación de células T autorreactivas en el timo, causando enfermedades autoinmunes<sup>184</sup>. A pesar de esta ambigüedad en la interpretación funcional de este polimorfismo, un estudio reciente confirma funcionalmente la hipótesis en la que la pérdida de función de LYP conllevaría una disminución en la regulación negativa de los linfocitos T y células dendríticas desencadenando una respuesta autoinmune. Esta pérdida de función vendría explicada por que la proteína LYP con el alelo de riesgo que codifica para un triptófano en la

posición 620, tiene una mayor susceptibilidad a la degradación por calpaína y proteosoma comparado con el alelo salvaje<sup>178</sup>. Por último, otro trabajo reciente ha identificado un nuevo polimorfismo de *PTPN22* asociado al LES, Arg263Gln localizado en el dominio catalítico, que también afectaría a la función fosfatasa. En este artículo, el alelo menor confiere protección frente a LES y provoca una reducción de la actividad fosfatasa de LYP<sup>185</sup>.

### 3.8. *PDCDI*:

El gen *PDCDI*, codifica para un receptor inmuno-inhibidor perteneciente a la familia B7/CD28 (**Figura 6**). Como ya se ha referido anteriormente, este locus fue uno de los primeros estudiados en relación al LES mediante estudios de ligamiento. El SNP (rs11568821) más asociado con la susceptibilidad del LES se denominó PD1.3 y está localizado en el cuarto intrón. Con respecto a su asociación a la susceptibilidad al LES, este *locus* también ha mostrado resultados controvertidos. En el primer estudio descrito anteriormente, el alelo de riesgo A de PD1.3 se encontró asociado a la susceptibilidad al LES en poblaciones suecas, euro-americanas y mejicanas<sup>186</sup>. Sin embargo, esta asociación no se confirmó en un estudio de asociación caso-control realizado en población española<sup>75</sup>. Además, se descubrió un sesgo geográfico en la frecuencia del alelo de riesgo A de PD1.3 a lo largo de Europa. Este alelo aumenta su frecuencia desde el Noreste hacia el Suroeste de Europa. Este sesgo solamente se encontró en los controles y no en los pacientes con LES<sup>187</sup>. Por otro lado, este gen no se ha encontrado asociado en ninguno de los GWAS realizados hasta la fecha tanto en población europea como en población asiática<sup>97-102</sup>. Finalmente, en un estudio reciente con población afro-americana, tampoco se ha confirmado la asociación de PD1.3 con la susceptibilidad al LES<sup>116</sup>.

Por otro lado, este polimorfismo PD1.3 se ha testado funcionalmente debido a su localización en una región con posible función de *enhancer*. En un primer estudio, mediante ensayos EMSA (*electrophoretic mobility shift assay*), se observó unión alélica diferencial del factor de transcripción RUNX1, relacionado con la diferenciación hematopoyética<sup>186</sup>. Este factor pertenece a una familia de factores de transcripción nucleares, CBF (*core binding factor*) teniendo como función principal la regulación de la expresión de genes implicados en la hematopoyesis y en la diferenciación de células mieloides y linfoides. Sin embargo, otro estudio evidenció la falta de funcionalidad de PD1.3 mediante la realización de estudios de gen *reporter* y EMSA. Para ello, en este estudio se evaluó el hipotético papel funcional de PD1.3 y el posible papel de la región como *enhancer* mediante la realización de estudios de gen *reporter* y EMSA<sup>76</sup>. En primer lugar, los resultados del EMSA confirmaron la unión preferencial del alelo

G. Por otro lado, utilizando los ensayos gen *reporter*, se analizaron los efectos transcripcionales de construcciones plasmídicas que incluían el intrón 4 de *PDCDI*. Los resultados no mostraron evidencias de que el intrón 4 funcione como *enhancer*. Además, tampoco se observó un efecto diferencial en la expresión génica a partir de los dos alelos de PD1.3.

### **3.9. *TNFAIP3* y región intergénica 6q23. *TNIP1*:**

El gen *TNFAIP3* codifica para la proteína A20 la cual juega un papel importante en la modulación de funciones celulares como la activación celular, la señalización de citoquinas y la apoptosis, actuando como molécula anti-inflamatoria<sup>188</sup>. Funcionalmente la proteína A20 consiste en un enzima que ubiquitina proteínas de la ruta pro-inflamatoria de NFκB, desencadenada por TLRs, el receptor de IL1 (IL1R), el receptor de TNF (TNFR) y NOD2<sup>189, 190</sup>. Las proteínas diana de A20 son proteínas adaptadoras como: TRAF6, RIP1, RIP2 e IKKγ/NEMO. Estas proteínas adaptadoras son ubiquitinadas lo cual conlleva una posterior degradación proteosómica<sup>190</sup> (**Figura 6**).

Este gen se mostró inicialmente asociado en uno de los GWAS realizados en LES con europeos, seguido de los dos GWAS en población asiática<sup>97, 101, 102</sup>. En el estudio de Graham R.R. et al.<sup>97</sup>, se postuló que la asociación del gen *TNFAIP3* con la susceptibilidad al LES vendría explicada por el efecto de un haplotipo de riesgo definido por los alelos menores de una serie de SNPs de baja frecuencia altamente correlacionados ( $r^2 > 0,98$ ): rs104999197, rs5029939, rs2230926 y rs7749323. Dentro de este haplotipo, está el SNP rs2230926 no sinónimo localizado en el exón 3 del gen *TNFAIP3*. Este polimorfismo da lugar a un cambio Phe127Cys de la proteína A20. Este nsSNP rs2230926 también se encontró asociado al LES de manera independiente en un estudio de mapeo fino del gen *TNFAIP3* y regiones adyacentes<sup>191</sup>. En este estudio se testó su funcionalidad comparando las diferencias en la habilidad de inhibir la producción de TNF mediada por NFκβ. Este análisis reveló que el alelo menor que da lugar a la proteína A20 con Cys127, presenta una ligera deficiencia a la hora de inhibir la expresión de TNF mediada por NFκβ. Pero esta diferencia no es significativa cuando la comparamos con la capacidad de inhibición de la proteína A20 codificada por el alelo mayor Phe127. Esta actividad anti-inflamatoria reducida de la proteína A20, podría explicar la excesiva repuesta celular al TNF en pacientes con LES. Aún así la funcionalidad de este SNP no está del todo clara. El meta-análisis de ambos estudios y resultados de imputación permitieron definir un haplotipo de riesgo que abarca 109 kb, en el que es difícil determinar si el alelo funcional es el nsSNP rs2230926 u otro de los 14 SNPs que constituyen el haplotipo<sup>192</sup>. Por último, en un reciente

estudio de mapeo fino del gen *TNFAIP3* se confirma la asociación del nsSNP rs2230926 en población con ascendencia europea y asiática<sup>193</sup>. Sin embargo, ninguno de los 127 SNPs que cubren el gen *TNFAIP3* y regiones adyacentes analizados en este estudio, muestran asociación significativa al LES en las poblaciones con ascendencia africana e hispana. Un resultado importante de este estudio consiste en que se define más restringidamente el haplotipo de riesgo al LES. Este haplotipo se extendería unas 16,3 Kb desde el SNP rs7752903 al SNP rs61117627. En él estarían involucrados no sólo el nsSNP rs2230926 sino que también un nuevo polimorfismo dinucleotídico TT>A con asociación independiente cuyo alelo raro define al haplotipo de riesgo al LES descrito anteriormente. Este polimorfismo dinucleotídico TT>A está localizado en una región de alta conservación a 42 Kb *downstream* del promotor del gen *TNFAIP3* en el cromosoma 6 posición 138,271,733. Por último, se analizó funcionalmente este polimorfismo dinucleotídico TT>A y no sólo se confirmó mediante EMSA la unión alélica diferencial del factor de transcripción NFκβ, sino que también se mostraron diferencias de expresión significativas de la proteína A20 usando los datos de expresión de las colecciones de líneas linfoblastoides del proyecto HapMap.

El gen *TNFAIP3* y la región intergénica *6q23* son dos *loci* vecinos que se encuentran en bloques haplotípicos distintos y que están asociados a varias enfermedades autoinmunes incluyendo AR<sup>194-196</sup> y LES<sup>97, 197</sup>. El SNP rs6920220 de riesgo que mostraba una asociación más fuerte con AR, está en la región intergénica *6q23*. Con respecto al LES, este SNP rs6920220 mostró una asociación modesta en el GWA de Graham y col.<sup>97</sup>. Sin embargo, otro estudio de replicación a gran escala confirmó la asociación de este SNP rs6920220 con la susceptibilidad al LES<sup>105</sup>. Finalmente, esta asociación se confirmó en otro estudio independiente en el que se encontraron asociados dos *tag* SNPs del rs6920220 en base a los datos del proyecto HapMap. De forma similar, el SNP protector para AR en la región intergénica (rs10499194) se vio asociado con LES en este estudio<sup>197</sup>. Otra de las señales de asociación que se encontró en esta región cromosómica es el SNP rs6922466 localizado en la región promotora de un gen implicado en apoptosis, *PERP* (*p53 apoptosis effector related to PMP22*), localizado a unas 200 kb 3' de *TNFAIP3*<sup>197</sup>. Sin embargo, esta asociación no ha sido descrita en ningún estudio adicional.

Por último en este apartado, introducir brevemente al gen *TNIP1* que codifica para una proteína que interacciona con TNFAIP3<sup>198</sup>; aunque su papel en la regulación de esta proteína se desconoce (**Figura 6**). Este *locus* mostró asociación al LES en poblaciones con ascendencia tanto europea como asiática<sup>101, 105</sup>. El SNP asociado, rs7708392, se localiza en un intrón del gen.

### 3.10. *ITGAM*:

El gen *ITGAM* codifica para integrina  $\alpha$ -M que dimeriza con la integrina  $\beta$ -2 (ITGB2 o CD18), formando el receptor del complemento 3 (C3b, también conocido como CD11b o Mac-1). Este complejo proteico se expresa en la superficie celular de macrófagos, monocitos, DCs y neutrófilos. A su vez el C3b interacciona con un amplio rango de ligandos, que incluyen el producto de activación del complemento 3 (C3bi), el fibrinógeno y la molécula 1 de adhesión intracelular (ICAM1). A partir de estas uniones, se llevan a cabo funciones tan variadas como: adhesión y trans migración de los leucocitos a través del endotelio, la activación de neutrófilos y monocitos, la apoptosis de neutrófilos y la eliminación de ICs mediante fagocitosis<sup>199-201</sup>. Esta última función es defectuosa en los pacientes con LES en los que se observan fallos en la eliminación de ICs, lo cual propone que polimorfismos presentes en *ITGAM* puedan contribuir con la susceptibilidad al LES (**Figura 6**).

El gen *ITGAM* se encuentra localizado en el locus *16p11* previamente asociado al LES en varios estudios de ligamiento con población de ascendencia europea<sup>69, 70</sup>. Además, esta asociación se ha confirmado en dos de los cuatro GWAS en población con ascendencia europea realizados para LES hasta la fecha<sup>98, 99</sup>. Un estudio de mapeo fino del gen encontró que el SNP más fuertemente asociado es un nsSNP rs1143679 (Arg77His) localizado en el exón 3 de *ITGAM*<sup>202</sup>. El papel funcional de este nsSNP rs1143679 todavía se desconoce, aunque se ha sugerido que podría afectar a la conformación y funcionalidad de la proteína alterando así la interacción de *ITGAM* con sus ligandos. La asociación de *ITGAM* (nsSNP rs1143679) ha sido fuertemente confirmada en un meta-análisis de nueve cohortes independientes de poblaciones con ascendencia europea, africana, hispana y asiática. Esta asociación se confirmó en todas las poblaciones exceptuando la población con ascendencia asiática en la cual este polimorfismo es prácticamente monomórfico, constituyendo un claro ejemplo de riesgo de LES específico de población<sup>203</sup>. Sin embargo, en otro estudio independiente, tanto nsSNP rs1143679 como otros polimorfismos dentro del gen *ITGAM*, han mostrado asociaciones independientes con una población de ancestralidad asiática<sup>204</sup>. Por otro lado, cuatro nuevos estudios independientes de los cuales dos de ellos están realizados en población europea, otro en población afro-americana y el último canadiense, han confirmado sólidamente la asociación de este locus *ITGAM* con la susceptibilidad al LES<sup>116, 158, 159</sup>.



### 3.11. C8orf13-BLK:

El gen *C8orf13*, también denominado *FAM167A* (*family with sequence similarity 167 member A*), se expresa de forma ubicua pero que no tiene función conocida. Por otro lado, el gen *BLK* se sabe que codifica una quinasa de tirosina de la familia Src, pero su función todavía no ha sido bien definida. En ratones, se ha sugerido que estaría implicada en la maduración de los linfocitos B, concretamente en la transición de células pro-B a pre-B<sup>205, 206</sup>. Sin embargo, en humanos se ha sugerido que BLK podría actuar en la señalización del receptor de células B (BCR) y de NFκB, influenciando no sólo en el desarrollo de células B sino que también en mecanismos de tolerancia y autoinmunidad<sup>207</sup> (**Figura 6**). Recientemente, se ha establecido una interacción epistática entre los genes *BLK* y *BANK1* en una población del norte de Europa. Además de la interacción epistática, también se observó una co-localización de ambas moléculas en la línea celular Daudi y en células B *naïve*<sup>208</sup>. La alteración de estas dos moléculas, desregularía la señalización de las células B ya que ambas moléculas están involucradas en la señalización del BCR y el control de desarrollo de las células B.

La asociación de esta región *C8orf13-BLK* radica en el estudio GWA de Hom G. et al. en el que se encontró un SNP rs13277113 localizado en la región promotora compartida por ambos genes asociado con la susceptibilidad al LES. Funcionalmente, se observó que el alelo A de riesgo del SNP rs13277113 está asociado con expresión reducida de *BLK* e incrementada de *C8orf13* en paciente con LES<sup>99</sup>. La asociación de este *locus BLK* con la susceptibilidad al LES también se confirmó en otros dos GWAS, además de un estudio de replicación independientes realizados todos ellos con poblaciones de ascendencia europea<sup>97, 98, 105</sup>. Además, esta señal se confirma en un meta-análisis alcanzando una significación a nivel de genoma completo<sup>114</sup>. Finalmente, otros estudios de replicación independientes adicionales confirmaron esta asociación en diferentes poblaciones con ascendencia europea y afro-americana<sup>116, 158, 159, 209</sup>. Con respecto a su asociación en poblaciones asiáticas, este locus ha resultado asociado en los dos GWAS realizados en estas poblaciones hasta la fecha<sup>101, 102</sup>. Esta asociación se ha replicado en un estudio posterior con población procedente de China<sup>210</sup>.

### 3.12. BANK1-LYN:

El gen *BANK1* codifica para una proteína adaptadora de células B que media la interacción entre la quinasa de tirosina LYN (*v-yes-1 Yamaguchi sarcoma viral-related oncogene homolog*) y el canal de calcio IP<sub>3</sub>R (*inositol 1,4,5-triphosphate receptor*), facilitando la liberación de

calcio intracelular, proceso que es importante en la activación de células B<sup>211</sup>. Por su parte, la quinasa de tirosina LYN regula de manera dual las células B. Por un lado las activa mediante la fosforilación del receptor BCR asociado a las Igαβ, y por el otro, media la inhibición de las células B mediante la fosforilación del receptor CD22<sup>49</sup>. En conjunto, estas funciones sugirieren que las variantes asociadas con LES podrían aumentar el riesgo a sufrir la enfermedad al alterar el umbral de activación de las células B (**Figura 6**).

El primer estudio que descubrió la asociación de *BANK1* con LES fue un estudio GWA de baja densidad<sup>100</sup>. En este estudio se definieron 3 variantes funcionales pero con efectos no independientes. El SNP rs17266594 se define como un polimorfismo *branch point* localizado antes del exón 2 mostró la mayor asociación. El alelo de riesgo de este SNP da lugar a una isoforma corta sin el exón 2 lo cual podría afectar a su unión con otras proteínas. La asociación de este SNP rs17266594 se confirmó en un estudio independiente en población europea<sup>158</sup>. También se encontró otro nsSNP rs10516487 (Ala383Thr) localizado a 153 pb del anterior y en alto LD con él ( $r^2 = 0,9$ ). La asociación de este último polimorfismo con la susceptibilidad al LES, se confirmó posteriormente en tres estudios de replicación independientes en poblaciones con ascendencia europea y afro-americana<sup>105, 116, 159</sup>. Sin embargo, en un meta-análisis con población de ascendencia europea, este SNP rs10516487 no presentó asociación con la susceptibilidad al LES<sup>114</sup>. El último polimorfismo asociado, es el SNP rs3733197 localizado en el exón 7, concretamente en el dominio anquirina de la proteína. Este SNP rs3733197 se considera un efecto independiente ya que no mostró LD con ninguno de los otros 2 SNPs ( $r^2 < 0.4$ ). La repercusión funcional de estos polimorfismos en la patogenia del LES no se ha descrito exactamente, pero éstos podrían influir en la sobre-activación del BCR y por consiguiente en la hiperactividad de las células B característica del LES. En relación a la asociación de este *locus BANK1* con el LES en población con ascendencia asiática, uno de los dos GWAS realizados mostró asociación de dos polimorfismos, SNP rs4522865 y SNP rs10516487 ambos sin llegar al nivel de significación de genoma completo<sup>102</sup>.

Las evidencias de asociación del gen *LYN* con el LES son menos numerosas. La primera evidencia de asociación se encontró en uno de los GWAS de LES con población de ascendencia europea, en el que encontraron el SNP rs2667978 con una significación a nivel de genoma completo<sup>98</sup>. Sin embargo, la asociación de este gen con la susceptibilidad al LES no se confirmó en varios estudios de replicación independientes en población con ascendencia europea<sup>105, 159, 209</sup>. Recientemente, se ha realizado un estudio de mapeado fino para este locus en poblaciones euro-americana, afro-americana y asiática. La asociación más fuerte encontrada en este estudio se sitúa en el SNP rs6983130, y sólo se ha demostrado en mujeres con ascendencia europea. La funcionalidad de este polimorfismo aún no ha sido descrita. Por otro lado, el SNP rs2667978

descrito anteriormente, muestra una asociación más débil en el mismo grupo de mujeres con ascendencia europea ( $P = 0,016$ ). Sorprendentemente, ninguno de los 90 polimorfismos analizados en este estudio, mostraron asociación con la susceptibilidad al LES en las poblaciones con ascendencia afro-americana o asiática<sup>212</sup>.

### 3.13. IRAK1-MECP2:

Los genes *MECP2* (*methyl CpG binding protein 2*) e *IRAK1* (*Interleukin-1 receptor-associated kinase 1*) están localizados en el mismo bloque de LD, concretamente en la región cromosómica *Xq28*. El gen *MECP2* codifica para una proteína que se une específicamente a ADN metilado. Esta proteína se unirá a su vez a deacetilasas de histonas (HDAC) y factores de transcripción como por ejemplo CREB1 (*cAMP responsive element binding protein 1*), promoviendo la inhibición o activación de la expresión del gen diana cuya expresión es sensible al estado de metilación del ADN<sup>213, 214</sup>. A su vez, se ha observado que el haplotipo de riesgo de *MECP2* estaba correlacionado con la sobre-expresión de genes sensibles a metilación en células B, la mayoría de ellos regulados por IFN<sup>215</sup>. Por otro lado, el gen *IRAK1* codifica para una quinasa de serina-treonina implicada en la cascada de señalización del receptor de la interleuquina 1 (IL1R). IRAK1 se une y activa a la proteína activadora/adaptadora TRAF6 que regula diversas rutas tanto en la respuesta inmune innata como adquirida. Se conoce que IRAK1 está involucrado en la inducción de la expresión de IFN $\alpha$  y IFN $\gamma$  además de ser un regulador de la ruta de señalización del NF $\kappa$ B<sup>216-218</sup>. Además, en modelos murinos de LES deficientes para *IRAK1* se observó que disminuían los títulos de auto-anticuerpos IgG y IgM, la activación generalizada de los linfocitos y la afectación renal<sup>219</sup>. Estas funciones están relacionadas con la patogenia del LES lo que llevó a sugerir que tanto *MECP2* como *IRAK1* contribuirían a la “firma del IFN” característica de esta enfermedad (**Figura 6**).

EN un primer estudio de asociación basado en la hipótesis de gen candidato, se evaluó la asociación del *locus MECP2* con la susceptibilidad al LES analizando un total de 21 SNPs en dos poblaciones, una con ascendencia europea y otra con ascendencia asiática. En este estudio se identificó al SNP rs17435, en el intrón 2, como el más fuertemente asociado tanto en población europeo-americana como en coreana. Además, se estudió la distribución haplotípica, así se encontró un haplotipo de susceptibilidad y otro de protección definidos por diferentes combinaciones alélicas del bloque de alto LD ( $r^2 > 0,8$ ) compuesto por los SNPs asociados: rs17435, rs1734791, rs1734792 y rs1734787<sup>220</sup>. De igual manera, tanto la asociación del *locus* como la misma distribución haplotípica se confirmó en un estudio independiente en el que se

utilizó una colección de muestras con ascendencia europea<sup>221</sup>. Posteriormente, la asociación de SNP rs17435, fue replicada en otro estudio de asociación con una colección de muestras europeas independientes<sup>158</sup>. Además, la asociación de la región *IRAK1-MECP2* (SNP rs2269368) se confirmó en un estudio de replicación a gran escala y en otro meta-análisis<sup>105, 114</sup>.

Por otra parte, el primer estudio centrado en *IRAK1* consistió en un análisis TDT de 9412 SNPs incluyendo 251 familias con un individuo afecto pero de LES infantil. El polimorfismo asociado a la susceptibilidad al LES infantil en este estudio es el nsSNP rs10127175 (Ser203Cys)<sup>222</sup>. Posteriormente, la asociación del *locus IRAK1* fue confirmada tanto en LES infantil como adulto, al genotipar 13 SNPs de la región en unas 10000 muestras de 4 grupos étnicos diferentes (europeos, asiáticos, africanos e hispanos). Los SNPs más asociados están en un bloque de alto LD, el cual se extiende desde el intrón 10 al intrón 13, que también define un haplotipo de riesgo y otro de protección<sup>223</sup>.

Como ya se ha señalado, *MECP2* e *IRAK1* se encuentran en un mismo bloque haplotípico. Sin embargo, estudios independientes mostraron de forma convincente asociación con *MECP2* o con *IRAK1*, pero ninguno de ellos ha analizado los dos genes. Por lo tanto, no es posible saber si la asociación se debe a uno o a otro. En los estudios para ambos genes, es posible que estén describiendo la misma asociación ya que los resultados en la presencia de dos efectos (susceptibilidad y protección) son muy similares. Por último, estos genes serían un ejemplo claro de la explicación del sesgo que existe entre hombres y mujeres con respecto a la prevalencia del LES. Por lo tanto, la dosis de alelos de riesgo de los genes *IRAK1* y *MECP2* podría contribuir a la mayor prevalencia del LES en mujeres, debido a su localización en el cromosoma X<sup>223, 224</sup>.

### 3.14. *TNFSF4*:

El gen *TNFSF4* (*TNF superfamily member 4*), también conocido como *OX40L*, codifica para el ligando del receptor *TNFSF4R* (CD134). Tanto el ligando como su receptor se expresan en células presentadoras de antígenos (CPAs), células endoteliales y células T CD4+ activadas. Esto sugiere que la interacción ligando-receptor, induce la producción de señales co-estimuladoras involucradas en la activación de las células T. Concretamente, esta cascada de señalización inhibe la expresión de IL10 a partir de las células T reguladoras CD4+ de tipo 1, pero por otro lado, induce la activación y diferenciación de células B así como la expresión de IL17<sup>225, 226</sup>. Para explicar el posible papel de este gen en la patogenia del LES, se hipotetizó que

altos niveles de este ligando llevarían a una elevada estimulación de las células T CD4+, aumentando así las interacciones célula T-CPAs y promoviendo la desestabilización de la tolerancia periférica (**Figura 6**).

La relación de este gen con LES parte de un GWAS en el que se encontró una fuerte señal de asociación con la susceptibilidad definida por un SNP rs10798269 localizado en el *locus* *Iq25.1*<sup>98</sup>. El único gen descrito en esa zona se encuentra a unas 133 kb *downstream*, consistiendo en el gen candidato *TNFSF4*. Posteriormente, otros cuatro estudios independientes replicaron y confirmaron la asociación de este *locus* con el LES en poblaciones independientes de ascendencia europea<sup>105, 114, 158, 159, 227</sup>. En un estudio de asociación no sólo caso-control sino que también basado en 472 familias europeas, se realizó un mapeado fino del gen *TNFSF4* y su receptor *TNFSF4R* (CD134). En este estudio se determinó la existencia de dos haplotipos definidos por SNPs de la región *upstream* del gen, uno de riesgo y otro de protección. Concretamente, el haplotipo de riesgo está relacionado con altos niveles de expresión de *TNFSF4*<sup>228</sup>. Finalmente, la asociación de este *locus* con el LES también se replicó en otras poblaciones con ascendencia afro-americana y asiática<sup>101, 102, 116</sup>.

### 3.15. *PHRF1-IRF7*:

El gen *PHRF1* (*PHD and RING finger domain-containing protein 1*) también conocido por *KIAA1542*, se encuentra en un *locus* homólogo al gen codificante de un factor de elongador. El SNP asociado con LES (rs4963128) se analizó en varios estudios independientes en población con ascendencia europea confirmando así su papel en la susceptibilidad al LES<sup>98, 105, 114, 158</sup>. Un estudio posterior con población afro-americana también confirmó esta asociación<sup>116</sup>. Sin embargo, un estudio reciente basado en familias con una población canadiense no consiguió replicar este resultado<sup>159</sup>. Se desconoce la función de la proteína que codifica y se ha sugerido que el efecto de la asociación con la susceptibilidad al LES observado en este *locus*, puede deberse a su proximidad con el gen *IRF7* (23 Kb *downstream*). Concretamente, el SNP rs709266 del gen *IRF7*, muestra un alto LD ( $r^2 > 0.9$ ) con el rs4963128<sup>98, 105</sup>. El factor de transcripción IRF7, ejerce una función importante en la señalización del IFN-I. Así, un estudio reciente muestra que el alelo de riesgo del SNP rs709266 está asociado con altos niveles de autoanticuerpos frente a ADNds e IFN $\alpha$  en pacientes con ascendencia europea<sup>229</sup>. Esta evidencia reforzaría la asociación del gen *IRF7* en lugar de *PHRF1* debido a su papel claro en la patogenia del LES (**Figura 6**).

### 3.16. ATG5-PRDM1-ETS1:

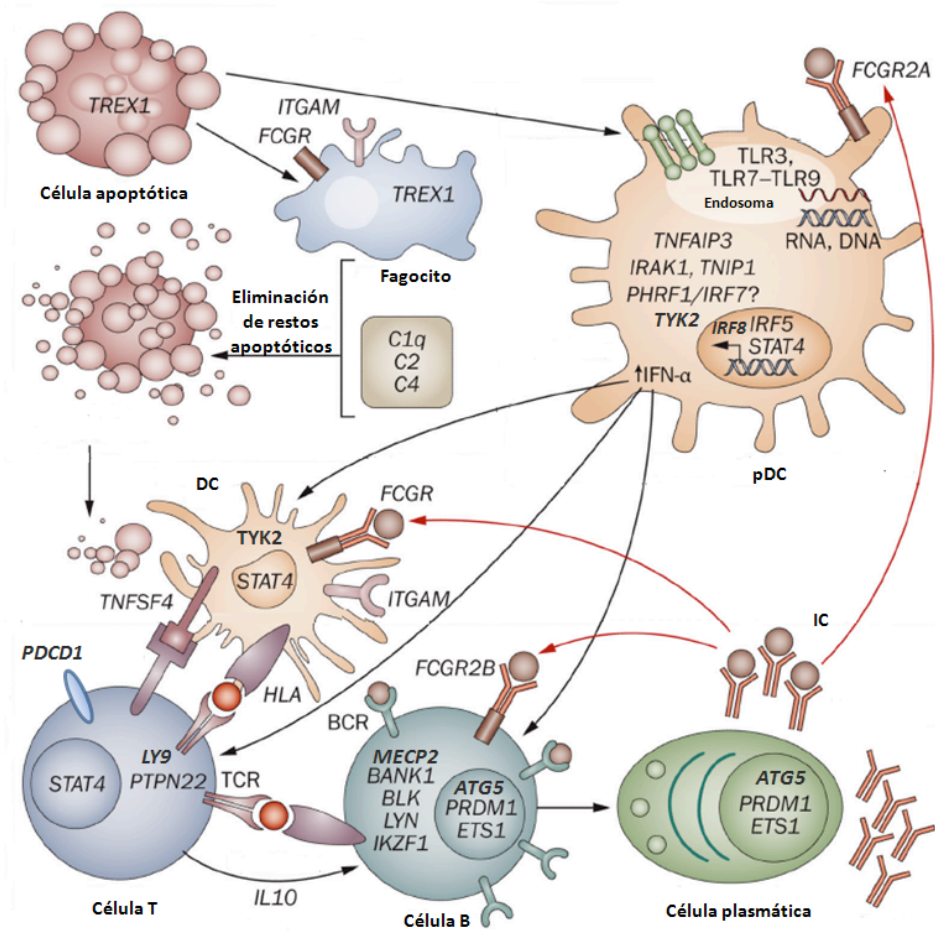
El gen *ATG5* codifica para una proteína implicada en la autofagia celular. Este proceso catabólico se define por la degradación del citoplasma y los orgánulos en exceso o aberrantes mediante lisosomas. Sin embargo, su papel en la patogenia del LES aún está por definir. Varios estudios han puesto de manifiesto que esta asociación con LES está localizada en el *locus PRDM1-ATG5*. Se han identificado varios SNPs involucrados en esta asociación localizados tanto en la región intergénica como dentro de cada *locus*<sup>98, 105, 114</sup>. Sin embargo, el bajo LD ( $r^2 < 0,1$ ) y los análisis condicionales, sugieren que son señales independientes. Por una parte, la asociación de *ATG5* no se ha replicado en dos estudios independientes con colecciones finlandesa y canadiense<sup>159, 209</sup>. Sin embargo, otra señal intergénica (SNP rs548234) ha mostrado asociación con LES en uno de los GWAS realizados en población asiática<sup>101</sup>. Con respecto a la funcionalidad del gen *PRDM1* (*PR domain zinc finger protein 1*), también conocido como *BLIMP1*, codifica para un factor de transcripción involucrado en la diferenciación de células T y B<sup>230</sup>.

Finalmente, se ha propuesto al *locus ETS1* (*protein C-ets-1*) como un nuevo factor genético involucrado en la patogenia del LES en población asiática<sup>101, 102</sup>. Este gen codifica para un factor de transcripción miembro de la familia ETS (*E-twenty six*) que regula negativamente la diferenciación de las células B y la Th17. Esta regulación se lleva a cabo mediante la inhibición del factor de transcripción PRDM1 expuesto anteriormente (**Figura 6**). Concretamente, el alelo de riesgo A del SNP rs1128334, localizado en la región promotora de *ETS1*, está asociado con la disminución de la expresión de ETS1 en PBMCs de sangre de controles<sup>102</sup>. Además, el ratón *knockout* para *ETS1* desarrolla un cuadro clínico de autoinmunidad caracterizado por altos títulos de autoanticuerpos y activación local del factor de complemento<sup>231</sup>.

### 3.17. LY9:

El gen *LY9*, también conocido como CD229, codifica para un antígeno leucocitario que se expresa en la superficie de monocitos, células NK y linfocitos T periféricos. Lo único que se conoce con respecto a su funcionalidad, es que participa en la adhesión entre células T y CPAs (**Figura 6**). *LY9* se encuentra en región de ligamiento al LES *Iq23*<sup>57</sup>. Este gen fue el que mostraba una asociación más fuerte de entre los estudiados en dicha región en un estudio TDT de 660 familias inglesas y 271 familias canadienses<sup>232</sup>. El polimorfismo más asociado fue el nsSNP rs509749 (Val602Met) presente en el exón 8 del gen *LY9*. Este nsSNP rs509749 cambia

la estructura proteica del dominio citoplasmático de la proteína LY9. Funcionalmente, el alelo de riesgo del nsSNP rs509749 distorsiona la población de células T incrementando el número de células T de memoria CD8+, además de disminuir la proporción de células T naïve CD4+ y células T activadas. Sin embargo, esta asociación no se replicó en otro estudio de asociación independiente en población europea<sup>158</sup>.



**Figura 6: Esquema de la interpretación funcional de las señales asociadas a la LES<sup>49</sup>.** Todos los genes están explicados en el texto. pDC: célula dendrítica plasmocitoide. DC: célula dendrítica. TCR: receptor de células T. BCR: receptor de células B. IC: inmuno-complejo.

#### 4. Nuevos genes asociados al LES.

Además de los genes comentados anteriormente, los GWAS y meta-análisis han permitido identificar muchos nuevos *loci* de asociación con el LES que no llegan al umbral de significación requerido por este tipo de estudios ( $P \leq 10^{-08}$ ). Sin embargo, estos *loci* presentan efectos importantes que podrían llegar a explicar parte de la “heredabilidad perdida” del LES. Por otro lado, muchos de estos nuevos genes no se ajustan a la definición de “gen candidato”, ya que no se ha definido una conexión directa entre su función y la patogenia de la enfermedad. Como valor aproximado, se han descrito alrededor de 30 nuevos *loci* de asociación con la susceptibilidad al LES. Dentro de este nuevo grupo de genes podemos destacar a *ICAI* (*islet cell autoantigen 1*), *XKR6* (*Kell blood group complex subunit-related family member 6*), *LYN* (*v-yes-1 Yamaguchi sarcoma viral-related oncogene homolog*), *UBE2L3* (*ubiquitin-conjugating enzyme E2L 3*), *SCUBE1*, *PXK*, *PTTG1*, *NMNAT2*, *JAZF1*, *UHRF1BP1*, *IL10* e *IRF8* en población europea<sup>98, 105</sup>. La mayoría estos genes sólo se han estudiado en uno o dos estudios independientes, por lo tanto, es necesaria la realización de más estudios de asociación para poder confirmar sus papel en la explicación de la susceptibilidad al LES. Sin embargo, hay algunos en los que se han realizaron varios estudios, como es el caso de *ICAI*, *XKR6*, *LYN*, *UHRF1BP1*, *UBE2L3* y *PXK*. Concretamente, *ICAI*, *LYN*, *XKR6* y *PXK* mostraron resultados de asociación ambiguos<sup>116, 158, 159, 209, 212</sup>. Sin embargo, en el caso del gen *UHRF1BP1*, se ha encontrado fuertemente asociado con la susceptibilidad al LES en dos poblaciones diferentes, europea y asiática<sup>105, 233</sup>. Hasta el momento se han descrito dos nsSNP (rs11755393, R454Q y rs13205210, M1098T) independientes dentro del gen *UHRF1BP1* como causantes de la asociación. Sin embargo, el papel funcional de este gen aún no es conocido, siendo lo único que se conoce que codifica para una proteína altamente conservada. En el caso de *UBE2L3*, todos los estudios realizados hasta la fecha confirman la asociación de este gen con la susceptibilidad al LES<sup>105, 114, 159</sup>. Además, se conoce que el alelo de riesgo T del SNP rs5754217 está fuertemente asociado con pacientes que presentan títulos frente a autoanticuerpos anti-La y anti-ADNds<sup>234, 235</sup>. Por lo tanto, esta podría ser una primera explicación que relaciona directamente su papel funcional con la patogenia característica del LES.

El papel de los genes *IL10* e *IRF8* en la patogenia del LES es mucho más consistente (**Figura 6**). *IL10* codifica para la interleuquina 10 la cual es una importante citoquina anti-inflamatoria<sup>236</sup>. Funcionalmente, se ha descubierto que el nivel de expresión de IL10 en células B y monocitos está correlacionado con el grado de actividad de la enfermedad en pacientes con LES<sup>237</sup>. Los resultados de asociación para este gen han sido inconsistentes en los estudios realizados hace unos años. Sin embargo, los últimos estudios parecen confirmar la asociación de este *locus* con la susceptibilidad al LES<sup>105, 238</sup>. En el caso del gen *IRF8*, éste codifica para el



factor regulador del interferón 8. Este factor de transcripción está presente en la ruta del IFN-I en la cual regula la expresión de múltiples citoquinas pro-inflamatorias. Por lo tanto, defectos en este gen podrían dar lugar a la “firma del interferón” característica de la patogenia del LES. Su asociación se ha replicado recientemente en un estudio de asociación independiente con 870 casos y 5551 controles ambos procedentes de Reino Unido llegando al nivel de significación de genoma completo<sup>156</sup>.

De la misma manera que en población europea, los GWAS realizados en población asiática también dieron lugar al descubrimiento de nuevos *loci* de asociación con la susceptibilidad al LES. Sin embargo, siendo muchos de ellos comunes en ambas poblaciones, hay otros que sólo están asociados en población asiática. Este fenómeno refleja una variación clara en el *background* genético de las diferentes etnias. Sin embargo, este fenómeno es una fuente de nuevas rutas de señalización a estudiar dentro de la patogenia del LES. Dentro de los genes identificados en población asiática, podemos destacar al gen *WDFY4* (*WDFY family member 4*) cuya asociación ha sido confirmada en los dos GWAS realizados en población asiática hasta la fecha<sup>101, 102</sup>. Sin embargo, su función concreta aún no se ha definido. Sin embargo, se conoce que codifica para una proteína altamente conservada y que se expresa principalmente en bazo, timo, nódulos linfáticos y amígdalas<sup>102</sup>.

## 5. Factor 5 regulador del IFN (*IRF5*).

La familia IRF (factores reguladores del interferón) comprende 9 miembros desde el IRF1 al IRF9. Estos factores comparten un dominio de unión constituido por un motivo *helix-turn-helix* de aproximadamente 120 aminoácidos. Esta región reconoce una secuencia consenso de DNA conocida como ISRE<sup>30</sup>. Esta secuencia consenso se encuentra en los promotores de múltiples genes que son activados por el IFN-I, dando lugar al patrón de expresión denominado “firma del IFN”. Algunos de estos genes son: *IFN $\alpha$* , *IFN $\beta$* , *IL-12* (interleuquina 12), *CXCL10* (*C-X-C motif chemokine 10 precursor*) e *IFIT1* (*interferon-induced protein with tetratricopeptide repeats*)<sup>239</sup>. Incluso el promotor de *IRF5* presenta una secuencia ISRE, por lo tanto, su expresión aumenta al activarse la ruta del IFN-I dando lugar a un fenómeno de retro-alimentación positiva<sup>240</sup>.

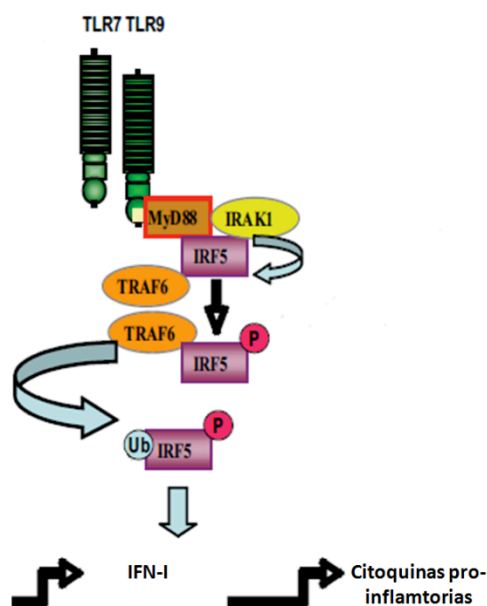
El gen *IRF5* se encuentra situado en el cromosoma 7q32. Este gen presenta una transcripción compleja ya que contiene 3 promotores conocidos y al menos 11 variantes de su ARN mensajero (V1-V11) debido a *splicing* alternativo<sup>241, 242</sup>. Este factor de transcripción se expresa en linfocitos B y células dendríticas de manera constitutiva, mientras que en monocitos

y macrófagos de manera inducida<sup>243-245</sup>. Su función principal es la inducción de la expresión de múltiples citoquinas pro-inflamatorias a través de la vía endosomal de los TLRs<sup>246</sup>. Sin embargo, *IRF5* también tiene un papel como gen supresor de tumores gracias a su efecto inductor de la apoptosis<sup>247-249</sup>. Se conoce poco sobre esta ruta de señalización, pero se ha descubierto que *p53*, otro gen supresor de tumores y regulador de la apoptosis, puede activar a *IRF5*. Sin embargo, las rutas de señalización apoptótica utilizadas por estos genes difieren<sup>250-252</sup>. También, se ha demostrado que *IRF5* inhibe el crecimiento de las células tumorales tanto *in vitro* como *in vivo*<sup>247</sup>.

En relación a estudios en modelos murinos, se ha descubierto que el ratón *knockout* para *IRF5* disminuye la expresión de múltiples citoquinas pro-inflamatorias como son TNF $\alpha$  e IL6 en DCs y macrófagos previamente tratados con ligandos de los TLRs<sup>253, 254</sup>. Sin embargo, a pesar de que el papel teórico de *IRF5* en la patogenia del LES está claro, no se ha descrito detalladamente en el desarrollo de la enfermedad. Un estudio reciente, ha analizado el papel de *IRF5 in vivo* en ratones MRL/lpr<sup>255</sup>. Esta línea de ratones desarrolla lupus murino, muy similar al LES humano, caracterizado por la producción de autoanticuerpos, hipergammaglobulinemia, glomerulonefritis, artritis, vasculitis, esplenomegalia y linfadenopatía<sup>256</sup>. El desarrollo de la enfermedad es debido a una mutación en el receptor de superficie celular Fas lo cual aumenta la apoptosis<sup>257, 258</sup>. En este estudio, cruzaron los ratones MRL/lpr con una línea de ratones *knockout* de *IRF5* obteniendo ratones MRL/lpr *IRF5*<sup>-/-</sup>. Analizando el fenotipo clínico de estos ratones MRL/lpr *IRF5*<sup>-/-</sup> comparándolos con los ratones MRL/lpr se observó una marcada disminución en la producción autoanticuerpos, glomerulonefritis leve y un aumento drástico en la supervivencia. Además, los ratones MRL/lpr *IRF5*<sup>-/-</sup> no sólo disminuyeron la producción de citoquinas pro-inflamatorias e IFN $\alpha$  por las DCs en respuesta a estimulación por TLRs, sino que también disminuyeron la producción de ICs en comparación con los ratones MRL/lpr. Este trabajo demuestra que el papel de *IRF5* en la patogenia del LES es importante e imprescindible para el desarrollo de al menos el lupus murino.

Un último estudio, ha postulado una nueva función del factor de transcripción IRF5 en la polarización de los macrófagos. Concretamente, la expresión de IRF5 determina la diferenciación de los macrófagos M1. Estos macrófagos son los encargados de mediar las respuestas inflamatorias en respuesta a patógenos y contribuyen a la destrucción del tejido. Por lo tanto, la alta expresión de *IRF5* en macrófagos M1, da lugar a el aumento de la expresión de múltiples citoquinas pro-inflamatorias como son IL12 subunidad p40 (IL12 p40), IL12 p35 y IL23 p19. Por el contrario, la expresión de *IRF5* en macrófagos M1 disminuye la expresión de la citoquina anti-inflamatoria IL10. Finalmente, se observó que estos macrófagos M1 polarizados eran funcionales ya que daban lugar a una respuesta Th1-Th17 correcta<sup>245</sup>.

IRF5 comparte la misma ruta de señalización con el factor de transcripción IRF7. Los factores IRF5 e IRF7 se unen a TLR7 y TLR9 los cuales se encuentran anclados en los endosomas y son activados mediante la internalización del ssRNA y dsDNA, respectivamente. La internalización de estos autoantígenos nucleares presentes en los ICs, está mediada por los receptores FCGR2 o FCGR3 los cuales los transportan desde la superficie celular hasta el endosoma<sup>259</sup>. Después de la activación de los TLRs se activa IRF5 el cual está anclado al endosoma mediante la proteína adaptadora MyD88. La primera molécula en actuar es la quinasa IRAK1 la cual activa a IRF5 fosforilándolo. Esta primera fase permite que la ubiquitin ligasa TRAF6 marque mediante ubiquitinización a IRF5 para su posterior translocación al núcleo. Este marcaje por ubiquitinización no degrada a IRF5<sup>260</sup>. El factor IRF5 marcado por ubiquitinización es translocado al núcleo donde se unirá a las secuencias ISRE presentes en los promotores de los genes de múltiples citoquinas pro-inflamatorias como son el  $\text{INF}\alpha/\beta$ ,  $\text{TNF}\alpha$ , IL-12 e IL6 probablemente en cooperación con  $\text{NF}\kappa\beta$ <sup>25, 244, 261</sup>. Por lo tanto, polimorfismos presentes en el gen *IRF5* podrían alterar la liberación de algunas de estas citoquinas favoreciendo así el desarrollo de LES<sup>53, 262</sup> (**Figura 7**).



**Figura 7: Ruta de señalización de IRF5<sup>263</sup>.** Todos los genes están explicados en el texto. P: grupo fosfato. Ub: ubiquitina.

En estudios previos, se ha analizado el patrón de genes expresados en linfocitos B humanos en relación con la señalización mediada por el factor de transcripción IRF5. Se demostró que un total de 657 genes (568 aumentados y 89 disminuidos) están regulados por IRF5. La regulación es directa mediante la unión de IRF5 a su promotor o indirecta mediante la vía IFN-I activada a su vez por IRF5<sup>264</sup>. Gracias a estos resultados, podemos ver la importancia de este factor en la regulación de la expresión de un amplio espectro de genes.

### 5.1. Asociación genética de *IRF5* con la susceptibilidad al LES:

Como ya se ha referido anteriormente, *IRF5* es uno de los genes asociados al LES con resultados ampliamente consolidados. En un primer estudio, se demostraron asociaciones de diferentes polimorfismos del *locus IRF5* con la susceptibilidad al LES<sup>155</sup>. Para ello se genotiparon 44 SNPs presentes en 13 genes de la vía IFN-I en una colección de 679 pacientes con LES procedentes de Suecia, Finlandia e Islandia, y un total de 1236 controles. Los datos obtenidos en este estudio muestran la clara asociación de 3 SNPs presentes en la región 5' del gen. Estos SNPs son el rs729302 ( $p=1,70 \times 10^{-3}$ ), el rs2004640 ( $p=2,4 \times 10^{-7}$ ) y el rs752637 ( $p=1,1 \times 10^{-3}$ ). Posteriormente, *IRF5* ha mostrado resultados de asociación significativos en cinco de los seis GWAS realizados en LES hasta la fecha no sólo en población con ascendencia europea sino que también con ascendencia asiática<sup>97-99, 101, 102</sup>. Además, *IRF5* también se ha encontrado asociado al LES en población con ascendencia afro-americana<sup>265, 266</sup>. Este resultado de asociación se ha reproducido y confirmado en múltiples estudios de asociación independientes<sup>105, 159, 209, 267, 268</sup>. Por otro lado, varios meta-análisis le han aportado consistencia a esta asociación en población europea y asiática<sup>114, 269, 270</sup>.

En población europea, se han identificado múltiples polimorfismos de susceptibilidad, incluyendo: un SNP cuyo alelo T crea un sitio de *splicing* en el exón 1B (rs2004640); un SNP cuyo alelo A crea un sitio de poli-adenilación en el extremo 3'UTR (rs10954213), resultando en una variante transcripcional más corta y de mayor estabilidad; un indel de 30 pares de bases en el exón 6, que afecta a un dominio de interacción de la proteína; y un indel de 5 pb en el promotor, que modifica un sitio de unión al factor de transcripción SP1 (*specificity protein 1*). Todos estos polimorfismos con repercusión funcional se han organizado en la distribución haplotípica de *IRF5* de tal manera que el haplotipo de riesgo para la enfermedad está constituido por el alelo T del SNP rs2004640, el alelo A del SNP rs10954213 y los alelos inserción de los dos indeles restantes. Este haplotipo se ha definido en varios estudios independientes y se ha asociado no sólo con una mayor expresión de *IRF5* sino que también de IFN $\alpha$ <sup>243, 268, 271-275</sup>. Sin embargo, en general, se ha confirmado que *IRF5* presenta una mayor expresión en pacientes con LES que en controles sin importar el haplotipo que presenten, indicándonos que tanto diferentes factores etiológicos como variantes genéticas influyen a la expresión de este gen en el LES<sup>266,</sup>

272.

### 5.1.1. SNP rs2004640: splice donor site en el exón 1B:

En un estudio posterior, se intentó reproducir la asociación del SNP rs2004640 con la susceptibilidad al LES. Para ello, se genotiparon cuatro colecciones de muestras procedentes de Estados Unidos, España, Suecia y Argentina, haciendo un total de 1661 casos y 2508 controles, y 470 familias multicaso<sup>243</sup>. Estos autores encontraron un aumento significativo de la frecuencia del alelo T del SNP rs2004640 en pacientes (60,4%) comparado con controles (51,5%) siendo  $p = 4,4 \times 10^{-16}$ . Además, el alelo T del SNP rs2004640 mostraba un exceso significativo de transmisión (219:153 T:U,  $p=0,0006$ ) utilizando el test TDT en las 470 familias estudiadas. Posteriormente, la asociación del SNP rs2004640 también fue confirmada en otro estudio con una colección de muestras independiente procedentes de Estados Unidos<sup>276</sup>. La frecuencia del alelo T del SNP rs2004640 fue significativamente mayor en casos que en controles (56% vs. 50%;  $p=0,008$ ) y la OR de los individuos portadores del alelo T fue 1,68 (95% I.C: 1,20-2,34;  $p=0,003$ ). Sin embargo, este no fue el único estudio que confirmó la asociación de este polimorfismo con la susceptibilidad al LES. Varios estudios posteriores confirman la asociación del SNP rs2004640 con el LES en colecciones con ascendencia europea<sup>271, 273, 274</sup>. Incluso, un estudio de asociación en el que se analizan 1488 pacientes con LES y 1466 controles de diferentes procedencias (México, Argentina, España, Italia y Alemania) confirman no sólo la asociación de SNP rs2004640 en conjunto si no que también en cada población individualmente<sup>268</sup>. Además, en este último estudio colocan al alelo T del SNP rs2004640 dentro del haplotipo de riesgo al LES. Recientemente, se ha confirmado la asociación de este SNP rs2004640 con el LES en otras dos poblaciones independientes procedentes de Canadá y Noruega con colecciones de tamaños más moderados que las anteriores<sup>159, 277</sup>. Además, en el estudio con la población Noruega, se encontró asociado este SNP rs2004640 con otras enfermedades autoinmunes como son la artritis idiopática juvenil (JIA) y más débilmente con la AR<sup>277</sup>. Sin embargo, la asociación del SNP rs2004640 con AR ya se había descrito sólidamente en un estudio de asociación previo con una colección de individuos españoles<sup>278</sup>.

El siguiente paso consistió en investigar la función del alelo T del SNP rs2004640. Este SNP se encuentra situado 2 pb *downstream* de la unión intrón 1-exón 1B. Su alelo T crea en este lugar una secuencia consenso GT la cual consiste en un *donor splice site*. Se ha demostrado que los individuos homocigotos para el alelo G del SNP rs2004640 expresan las isoformas de *IRF5* que contiene los exones 1A y 1C, pero no el exón 1B. Por otro lado, los individuos homocigotos para el alelo T expresan los transcritos que contienen el exón 1B además de otros transcritos que contienen el exón 1A y el exón 1C en menor medida. Estos exones constituyen la región 5'-UTR del gen y, por lo tanto, no se traducen<sup>243</sup>. Aún no se ha determinado si las distintas isoformas de *IRF5* tienen repercusión funcional. Pero se conoce que las isoformas del mRNA de

*IRF5* que contienen el exón 1A predominan sobre las isoformas que contienen los exones 1B, 1C y 1D<sup>274</sup>. Por lo que es dudoso que la presencia de la isoforma con el exón 1B tenga una importante repercusión funcional.

Además, se ha confirmado que los exones 1A y 1D presentan mayor expresión en pacientes con LES que en controles. Por el contrario, la expresión de los exones 1B y 1C tanto en pacientes con LES como en controles es muy baja o imperceptible<sup>272</sup>. En este mismo estudio también se confirmó que el haplotipo de riesgo para el LES contenía el alelo T del SNP rs2004640. Este haplotipo de riesgo presenta una mayor expresión de *IRF5* en PBMCs de pacientes con LES. Además, la expresión de todos los exones 1 fue mayor en los pacientes con el haplotipo de riesgo que en los pacientes con el haplotipo protector. Consecuentemente, el haplotipo de riesgo presenta una mayor expresión proteica de IRF5 en varios tipos celulares como son: monocitos, mDCs y células B. Por último, en este estudio se observó una tendencia no significativa hacia el aumento no sólo en la expresión de *IRF5* en PBMCs de pacientes con el alelo T del SNP rs2004640, sino que también en la producción proteica de este factor de transcripción en mDCs de pacientes con el alelo T del SNP rs2004640.

Por otro lado, un estudio reciente ha correlacionado el aumento de expresión de IFN $\alpha$  en suero de pacientes con LES con el haplotipo de riesgo que contiene el alelo T del SNP rs2004640 con respecto a los pacientes con LES portadores de haplotipos neutros y protectores. Sorprendentemente, este efecto es más notable en el subgrupo de pacientes con LES que presentan autoanticuerpos anti-RBP o anti-dsDNA<sup>275</sup>.

### **5.1.2. SNP rs10954213: señal de poly A:**

Diferentes grupos re-secuenciaron el gen *IRF5* y secuencias vecinas en busca de nuevos polimorfismos funcionales que explicasen su asociación con la susceptibilidad al LES. Se re-secuenciaron todos los exones e intrones del gen además de 1 Kb *upstream* del exón 1A en casos y controles procedentes de Suecia y Estados Unidos. Se encontraron 52 polimorfismos, de los cuales 26 ya habían sido identificados. Dos SNPs (rs10954213 y el rs10954214) están presentes en la zona 3'-UTR, y están en alto LD con el SNP rs2280714<sup>273</sup>. Este SNP, rs2280714, se había encontrado asociado con niveles altos de expresión de *IRF5* con independencia del LES<sup>279, 280</sup>. Debido a esto, se analizó la correlación entre estos 2 SNPs y el nivel de expresión de *IRF5*. El alelo A del SNP rs10954213 está significativamente correlacionado con una alta expresión de *IRF5*. También está asociado al riesgo de padecer LES ya que su frecuencia en casos (67,2%) es mayor que en controles (61,2%) siendo  $p = 9,1 \times 10^{-5}$ <sup>273</sup>. Estos resultados se verificaron en un estudio de asociación caso-control independiente con

un grupo de muestras procedentes de Argentina, España y Alemania (frecuencia alelo A rs10954213 en casos 66% vs. 60% en controles;  $p=0,0008$ )<sup>274</sup>. Además, en otro nuevo estudio del mismo grupo se analizan 1488 pacientes con LES y 1466 controles de diferentes procedencias (México, Argentina, España, Italia y Alemania) confirmando la asociación de SNP rs10954213 en conjunto pero no en cada población individualmente<sup>268</sup>.

El SNP rs10954213 se localiza en una señal de poly A (poli-adenilación) altamente conservada [AAT(A/G)AA] dentro de la zona 3'-UTR del gen *IRF5*. Esta señal de poly A funciona como lugar de unión para el complejo CPSF (*cleavage and polyadenylation specificity factor*). Durante la transcripción mediada por la RNA polimerasa II, el complejo CPSF se une a la señal de poly A (AATAAA), corta el mRNA entre 10-30pb *downstream* de la señal poly A y comienza la poliadenilación de los transcritos. Por lo tanto, la presencia del alelo A da lugar a un mRNA de menor tamaño ya que se produce la poli-adenilación 12 pb *downstream* de la señal poly A. Por otro lado, en presencia del alelo G no se produce poli-adenilación hasta la siguiente señal de poly A situada a 648 pb *downstream*. Mediante ensayos *Northern Blot* y *qPCR*, se verificó que las células homocigotas para el alelo A expresan principalmente la versión corta del mRNA de *IRF5*. Por el contrario, las células homocigotas para el alelo G expresan exclusivamente la versión larga del mRNA de *IRF5*. Estudios de *gen reporter* utilizando constructos con las dos posibles regiones 3'-UTR del gen seguido del gen de la  $\beta$ -globina confirmaron que las células homocigotas para el alelo A del SNP rs10954213 tienen un nivel de expresión 5 veces superior a las células homocigotas para el alelo G<sup>273</sup>. Estos resultados fueron confirmados en otros tres estudios. En uno de ellos se utilizó el *array* de expresión de *Affymetrix* y mRNAs de líneas celulares linfoblastoides<sup>271</sup>. En un segundo estudio, se usaron ensayos de *qPCR* con PMBCs estimuladas con IFN $\alpha$  y sin estimular<sup>274</sup>. En tercer lugar, se ha confirmado que la expresión de *IRF5* está aumentada en PMBCs de pacientes con LES con el alelo A del SNP rs10954213 y que esta contribución al aumento de la expresión es independiente y aditiva al aumento provocado por el alelo de riesgo del indel CGGGG que se explicará a continuación<sup>268</sup>.

En un estudio reciente, se ha estudiado la distribución haplotípica de *IRF5* colocando al alelo A del SNP rs10954213 dentro del haplotipo de riesgo. En este estudio también se confirmó que este haplotipo de riesgo aumentaba la expresión de *IRF5* en PMBCs de pacientes con LES. Consecuentemente, el haplotipo de riesgo presenta una mayor expresión proteica de *IRF5* en varios tipos celulares como son: monocitos, mDCs y células B. Por último, en este estudio se observó una tendencia no significativa hacia el aumento no sólo en la expresión de *IRF5* en PMBCs de pacientes con el alelo A del SNP rs10954213, sino que también en la

producción proteica de este factor de transcripción en mDCs y monocitos de pacientes con el alelo A del SNP rs10954213<sup>272</sup>.

Teniendo en cuenta todos los resultados podemos concluir que el alelo de riesgo A del SNP rs10954213 se transcribe como un mRNA con una cola de poly A corta lo cual le aporta mayor estabilidad. Esto da lugar a una expresión aumentada del mRNA corto que puede resultar en una excesiva producción de IFN-I mediada por IRF5.

### **5.1.3. Inserción/delección presente en el exón 6: región rica en prolina:**

El siguiente polimorfismo funcional que se ha encontrado en el gen *IRF5* consiste en una inserción/delección de 30 pb presente en el exón 6. Este polimorfismo da lugar a la inclusión o pérdida de 10 aminoácidos en una región rica en prolina<sup>273, 274</sup>. La unidad de repetición de este polimorfismo consiste en una secuencia de 14 nucleótidos (GCCGCCCACTCTGC). En el alelo delección esta unidad está repetida dos veces y en la inserción esta unidad se repite 4 veces. La presencia o ausencia de las repeticiones determina las isoformas de *IRF5* que se están expresando. Así los individuos homocigotos para la inserción (4 repeticiones) expresan las isoformas V5 ó V6, los individuos homocigotos para la delección (2 repeticiones) expresan las isoformas V1 ó V4. Todas las isoformas en cada genotipo tiene el mismo nivel de expresión<sup>274</sup>. A pesar de la repercusión funcional posible de este polimorfismo, este indel no se ha encontrado asociado a la susceptibilidad al LES analizada de manera independiente<sup>273, 274</sup>. Sin embargo, la inserción del exón 6 presenta asociación al LES en análisis de regresión logística condicionada con los otros dos polimorfismos funcionales explicados previamente (rs2004640 y rs10954213)<sup>273, 274</sup>.

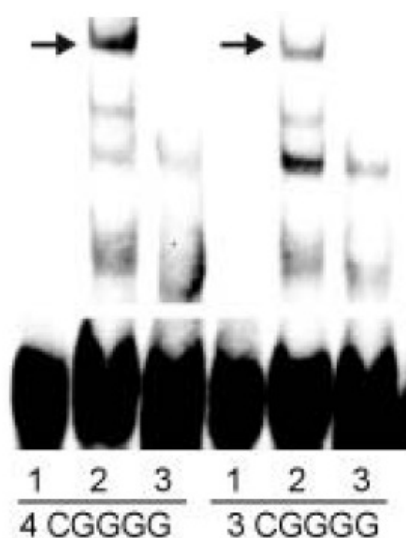
Un estudio reciente ha analizado el posible efecto funcional de este indel junto con otra variante presente dentro del exón 6 de *IRF5*. Esta nueva variante consiste en la pérdida de 48bp (16 aminoácidos) debido a la presencia de un sitio de *splicing* diferente al del inicio del exón 6. Este nuevo sitio de corte, dará lugar a una nueva variante de *splicing* con 16 aminoácidos menos en la proteína (SV-16). Se analizaron todas las posibles combinaciones entre estos dos polimorfismos. Concretamente, se observó que las variantes de *IRF5* que contienen una u otra inserción, no sólo disminuyen su capacidad en la inducción de apoptosis sino que también disminuían la translocación al núcleo de este factor de transcripción. En relación a la expresión de citoquinas pro-inflamatorias, sólo la variante con los 16 aminoácidos intactos pero con la delección de los 10 aminoácidos en el exón 6, aumenta la expresión de IL6. Por el contrario, solamente la variante con la delección tanto de los 10 aminoácidos como de los 16 aminoácidos disminuye la expresión de esta citoquina<sup>281</sup>.



#### 5.1.4. Inserción/delección CGGGG: unión al factor de transcripción SP1:

El mismo grupo de autores que identificó por primera vez la asociación de *IRF5* con LES ha proseguido el estudio de este gen centrándose en un polimorfismo de tipo indel presente en el promotor. Este polimorfismo está localizado 64 pb *upstream* del inicio del exón 1A de *IRF5*. La unidad de repetición de este polimorfismo consiste en una secuencia de 5 pb (CGGGG) que se repite 4 veces (inserción) o 3 veces (delección)<sup>273</sup>. Este polimorfismo está asociado al LES y a enfermedad inflamatoria intestinal<sup>282, 283</sup>. En cuanto al LES, se genotiparon 485 pacientes y 563 controles procedentes de Suecia<sup>283</sup>. Como resultado, la inserción 4xCGGGG presentó una frecuencia en casos (56%) superior que en controles (44%) dando lugar a una OR = 1,69 (95% I.C.: 1,42-2,02) siendo  $p = 4,6 \times 10^{-9}$ .

Ensayos EMSA mostraron que la inserción 4xCGGGG crea un lugar de unión adicional en la región promotora de *IRF5* para el factor de transcripción SP1<sup>282, 283</sup> (**Figura 8**). Además, ensayos *reporter* con mini-genes y análisis *immunoblot* muestran que las células homocigotas para la inserción 4xCGGGG presentan niveles de expresión de *IRF5* que contiene el exón 1A entre 2 y 10 veces superiores a las células homocigotas para la delección 3xCGGGG<sup>283</sup>.



**Figura 8:** Ensayo EMSA con oligonucleótidos de 37pb para la inserción (4xCGGGG) y delección (3xCGGGG) y extracto nuclear de células sanguíneas<sup>282</sup>. (1) sonda marcada; (2) sonda marcada y extracto nuclear; (3) sonda marcada y exceso de sonda sin marcar más extracto nuclear. La banda de unión específica de SP1 con las sondas para cada alelo está marcada con una flecha.

Por otro lado, mediante análisis de regresión logística condicionada, se vio que la asociación de *IRF5* al LES puede explicarse por dos efectos independientes definidos por el SNP rs10488631 y el indel CGGGG<sup>283</sup>. Incluso, en un estudio de asociación en el que se analizan 1488 pacientes con LES y 1466 controles de diferentes procedencias (México,

Argentina, España, Italia y Alemania) confirman no sólo la asociación del indel CGGGG en conjunto si no que también por cada población individualmente<sup>268</sup>.

Finalmente, el alelo inserción de este polimorfismo se ha colocado dentro del haplotipo de riesgo del gen *IRF5* al LES<sup>268, 272</sup>. Además, en sentido de los resultados previos, se observó una tendencia no significativa hacia el aumento no sólo en la expresión de *IRF5* en PBMCs de pacientes con el alelo inserción, sino que también en la producción proteica de este factor de transcripción en mDCs y monocitos de pacientes con el este alelo<sup>272</sup>. Por último, se ha confirmado que la expresión de *IRF5* está aumentada en PBMCs de pacientes con LES con el alelo inserción, y que esta contribución al aumento de la expresión es independiente y aditiva al aumento provocado por el alelo de riesgo A del SNP rs10954213<sup>268</sup>.

## 6. Estudio de la expresión génica: *microarrays* de expresión y secuenciación del ARN (RNA-seq).

A lo largo de estos años, se han desarrollado varias técnicas capaces de analizar la expresión génica a nivel de genoma completo. Estas técnicas estudian las variaciones en el transcriptoma. El transcriptoma se puede definir como el conjunto total y cantidad de transcritos presentes en una célula. Lógicamente, el transcriptoma varía dependiendo del momento concreto en el desarrollo de la célula, las condiciones fisiológicas y por último del tipo celular. Por lo tanto, el estudio del transcriptoma en condiciones patológicas nos puede informar sobre qué rutas de señalización pueden estar des-reguladas y por lo tanto, qué componentes pueden estar explicando una determinada enfermedad.

La primera metodología desarrollada para el estudio del transcriptoma son los *microarrays*. Concretamente, los *microarrays* de ADN son utilizados principalmente en estudio de la expresión génica. El principio general de esta técnica se basa en la hibridación de moléculas de ADN complementarias. Así, se diseñan *microarrays* con múltiples sondas homólogas a diferentes zonas cubriendo la expresión génica a nivel de genoma completo. El tamaño de las sondas varía de entre 25 a 60 nucleótidos. La metodología se resume en la extracción de la totalidad del ARN mensajero (ARNm) a partir de la muestra de estudio, seguida de la retro-transcripción del ARNm a ADN complementario (ADNc), marcaje de este ADNc con fluoróforos y, por último, la hibridación con el *microarray* de sondas. El ADNc se marca habitualmente con fluoróforos lo cuales emitirán una señal de fluorescencia relativa a la cantidad de sonda a las que se haya unido. Esta señal nos da un dato cuantitativo de la expresión de un gen o secuencia específica del genoma.

Un tipo de *microarray* muy útil en el estudio de enfermedades, son los *microarrays* de dos canales. En estos *microarrays* se analizan al mismo tiempo dos muestras con diferentes condiciones (por ejemplo, paciente y control). Esto se lleva a cabo mediante el marcaje de los dos ADNc de estudio con dos fluoróforos que emiten la fluorescencia a diferentes longitudes de onda. Los resultados de este tipo de experimentos nos informan sobre las intensidades relativas de cada fluoróforo frente al otro, de manera que podremos identificar qué genes están desregulados bajo las condiciones de nuestro estudio. Es importante destacar que estos estudios no son cuantitativos, a diferencia de los *microarrays* de un solo canal explicados anteriormente.

A pesar de su gran utilidad, los *microarrays* tienen una serie de limitaciones como son: alta dependencia con el conocimiento del genoma en el momento del diseño de las sondas, altos niveles de ruido de fondo debido a la hibridación cruzada y finalmente, un pequeño rango dinámico de valores debido a este alto ruido de fondo y a la temprana saturación de la señal.

Recientemente, se ha desarrollado una nueva tecnología de secuenciación de alto rendimiento (NGS, *next generation sequencing*). Este tipo de secuenciación es utilizada para el estudio del transcriptoma tecnológicamente denominado RNA-seq (secuenciación del ARN). La metodología del RNA-seq consiste en la extracción del ARNm a partir de la muestra de estudio, seguida de la retro-transcripción del ARNm a ADNc, este ADNc será fraccionado en fragmentos que serán amplificados y secuenciados mediante NGS. Gracias a este tipo de secuenciación, se consiguen múltiples secuencias de regiones de entre 30-400 pb dependiendo de la tecnología empleada. A estos fragmentos resultantes de la secuenciación se los denomina lecturas (*reads*). Por lo tanto, el número de *reads* por región nos informa de manera cuantitativa sobre la expresión de esa misma región.

La tecnología RNA-seq presenta múltiples avances en relación a los *microarrays*. El primero de ellos consiste en que RNA-seq no está condicionado al conocimiento del genoma sino que nos permite descubrir nuevas variantes de *splicing*. Otra ventaja de esta metodología consiste en que presenta un bajo, incluso casi inexistente, ruido de fondo ya que los *reads* se alinean de manera específica con una única región complementaria del genoma. Por último, esta técnica no presente ningún límite en la cuantificación de la expresión ya que los datos obtenidos son digitales. Por lo tanto, el rango dinámico de valores es mucho más amplio que el se obtiene utilizando *microarrays*. Teniendo en cuenta todas estas características, podemos decir RNA-seq es el primer método basado en secuenciación que nos permite estudiar el transcriptoma en su totalidad obteniendo una resolución a nivel de un pb y un resultado de expresión cuantitativo y digital<sup>284</sup>.

## 7. Heterogeneidad del LES.

### 7.1. Heterogeneidad clínica:

El LES es una enfermedad muy heterogénea desde el punto de vista clínico. Presenta 11 manifestaciones clínicas (criterios ACR), sin embargo, los pacientes no tienen porque presentar todas las manifestaciones clínicas al mismo tiempo, siendo la presencia de al menos 4 de ellas el criterio clínico para definir un paciente con LES. Esta definición de enfermedad ya nos lleva a una subclasificación de los pacientes con LES en base a sus manifestaciones clínicas. Por otro lado, el curso del LES se caracteriza por exacerbaciones y remisiones. Durante estas fases agudas de la enfermedad, el paciente puede presentar diferentes manifestaciones clínicas. Todo estos aspectos hacen que el LES sea clínicamente muy heterogéneo dificultando así su diagnóstico, tratamiento y por consiguiente su estudio.

Por otro lado, se ha demostrado que el componente genético no sólo afecta a la susceptibilidad de la enfermedad, sino que también a la forma clínica. Así varios estudios epidemiológicos, han observado que el riesgo a padecer manifestaciones severas de la enfermedad varía según el grupo étnico. En general, en poblaciones no caucásicas, los pacientes tienden a ser más jóvenes y a sufrir una enfermedad más grave y activa, con afectación renal, cardiovascular y alta producción de autoanticuerpos<sup>143, 285</sup>. Este fenómeno nos lleva a hipotetizar que el *background* genético de cada grupo étnico podría influenciar la prevalencia de los distintos fenotipos clínicos. Se define prevalencia como la proporción de individuos de un grupo o una población que presentan una característica clínica determinada en un momento o en un período determinado. Algún ejemplo de este fenómeno se puede observar dentro de manifestaciones clínicas concretas. Así los pacientes europeos muestran una mayor prevalencia de manifestaciones muco-cutáneas, en particular fotosensibilidad, comparado con otros grupos étnicos<sup>286-288</sup>. Sin embargo, los pacientes con ascendencia africana, presentan una mayor prevalencia de eritema discoide y menor prevalencia de eritema malar y fotosensibilidad comparado con pacientes europeos<sup>288-290</sup>. Por otro lado, tanto los pacientes con ascendencia africana, asiática e hispano-americana presentan una mayor prevalencia de nefritis comparado con los pacientes europeos<sup>291</sup>.

En el mismo sentido, se sabe que existen diferencias genéticas entre poblaciones del mismo continente, lo que sugiere que también podría existir una contribución de la ascendencia genética, dentro de una misma etnia, al desarrollo de las distintas manifestaciones clínicas del LES. Este es el caso de Europa dentro de la cual se define un fenómeno de subestructura que divide a las poblaciones en subgrupos en base a su *background* genético. Esta subestructura europea fue analizada en un estudio en el que se genotiparon alrededor de 5700 SNPs en un total

de 928 pacientes con ascendencia europea<sup>292</sup>. Mediante el análisis de componentes principales de los genotipos, se observó que existían dos componentes principales que definen los ejes geográficos norte-sur y este-oeste de Europa. Los SNPs que mejor definen las dos componentes principales se consideran marcadores informativos de ancestralidad (AIMs). Estos AIMs se caracterizan por mostrar grandes diferencias en frecuencias entre los subgrupos europeos<sup>292-295</sup>.

Gracias a la utilización de los AIMs, varios estudios previos han observado una correlación entre la prevalencia de algunos fenotipos clínicos del LES con la distribución de sus frecuencias genéticas<sup>296, 297</sup>. Por lo tanto, estos resultados nos indican que existen manifestaciones clínicas con diferencias en su prevalencia a lo largo de los dos ejes geográficos de Europa. Así, se ha encontrado una mayor prevalencia de manifestaciones muco-cutáneas (fotosensibilidad y eritema discoide) en pacientes con ascendencia del norte de Europa. Por otro lado, los pacientes que procedían del sur de Europa presentaban una prevalencia de producción de autoanticuerpos y de nefritis. Finalmente, los pacientes con origen en el oeste de Europa presentaban más serositis y producción de autoanticuerpos en comparación con los pacientes del este<sup>296, 297</sup>.

Finalmente, estos resultados sugieren que incluso pequeñas diferencias en la ascendencia genética podrían influenciar el desarrollo de manifestaciones específicas. Sin embargo, hay que destacar que estos estudios se realizaron con colecciones de muestras con ascendencia europea procedentes de Estados Unidos. Este hecho hace que haya un gran predominio de los pacientes con origen del norte de Europa, por lo que sería interesante abordar esta cuestión con muestras obtenidas en Europa y con una representación más equilibrada.

## **7.2. Heterogeneidad genética:**

Como ya hemos comentado en apartados previos, se han descrito más de 30 *loci* asociados con la susceptibilidad al LES. Estos resultados dejan clara la gran heterogeneidad presente en el componente genético de esta enfermedad. Sin embargo, todas estas señales no están asociadas en todas las etnias. Varios *loci* de asociación al LES como son *STAT4*, *IRF5*, *BLK*, *TNFSF4*, *TNFAIP3* y *BANK1* presentan resultados de asociación similares en diferentes grupos étnicos. Sin embargo, otros como *PTPN22*, *PXK*, *FCGR2A* y los genes de la región HLA, presentan resultados discordantes dependiendo la etnia estudiada<sup>97-100, 102, 116</sup>. Estas diferencias pueden deberse a varias causas como son: falta de polimorfismo en sus frecuencias en alguna etnia, diferencias del patrón de LD entre etnias o incluso diferencias en las interacciones con otros genes o factores ambientales entre las etnias. Apoyando esta hipótesis, en un estudio reciente se genotiparon 16 *loci* de asociación al LES en 804 pacientes con LES y 667 controles ambos con

ascendencia amerindia<sup>298</sup>. Al comparar los resultados de asociación en esta población con los resultados previos de diferentes etnias, se observó un aumento en el tamaño del efecto en los individuos con ancestralidad amerindia. Además, se observó la presencia de un aumento de 2,34 en el número de alelos de riesgo en los pacientes con ascendencia amerindia en comparación con las demás etnias estudiadas.

Por otro lado, también se ha planteado la hipótesis de que estas diferencias en el tamaño del efecto de las señales de asociación al LES no sólo se vean entre etnias diferentes sino que también dentro de la misma etnia. Esto lleva a pensar que dentro de Europa también encontremos estas diferencias debido al fenómeno de subestructura descrito en el apartado anterior. Esta hipótesis vendría reforzada por los resultados comentados previamente en relación a las diferencias de prevalencia de las variables clínicas a lo largo de Europa y por algún estudio como el del gen *PDCDI* el cual ha mostrado diferencias de asociación claras en varias colecciones de Europa<sup>75</sup>. Es importante destacar, que ambos resultados muestran variaciones a lo largo del eje geográfico europeo norte-sur, siendo este el eje principal para la definición de la subestructura europea. Todas estas evidencias en conjunto hacen interesante el estudio de la heterogeneidad genética del LES a lo largo de la subestructura europea.

Finalmente, también se ha descrito la existencia de asociaciones entre varios *loci* de asociación al LES con el aumento o disminución de la prevalencia de diferentes fenotipos clínicos. Este fenómeno tiene evidencias claras y contribuye a una posible causa de la heterogeneidad genética del LES. Varios estudios han corroborado este fenómeno. Por ejemplo, el gen *STAT4* está asociado con una mayor prevalencia de nefritis, edad temprana de inicio de la enfermedad y desorden neurológico. Sin embargo, tiene un efecto protector frente a la manifestación de úlceras orales<sup>164</sup>. Otro gen importante en el LES es *ITGAM*, el cual se ha asociado con una mayor prevalencia nefritis, desorden inmunológico y eritema discoide<sup>299</sup>.

Recientemente, se han publicado varios estudios en los que se analiza un espectro más amplio de genes en relación a las manifestaciones clínicas del LES. Así en estos estudios se analizaron entre 16 y 22 *loci* de asociación al LES en varias colecciones con ascendencia euro-americana, hispana, afro-americana y asiática<sup>172, 174</sup>. En estos estudios se confirma la asociación del gen *ITGAM* con el aumento de la prevalencia de nefritis y eritema discoide en el conjunto de todas las poblaciones; además de la asociación del gen *STAT4* con la protección frente a la presencia de úlceras orales de nuevo en el conjunto de todas las colecciones<sup>172</sup>. También se describen nuevas asociaciones presentes en todas las etnias estudiadas como son: la asociación del *FGR2A* con el aumento de la prevalencia del eritema malar, la asociación de *TNFSF4* con el aumento de prevalencia de la nefritis y por último, la asociación de *IL21* con el aumento de la prevalencia del desorden hematológico<sup>172</sup>. Por otro lado, ya en población europea, se confirma

la asociación del gen *STAT4* con la edad temprana de inicio de la enfermedad<sup>174</sup>. Además, se describen nuevas asociaciones presentes en población europea como son: la asociación de *HLA-DR3* con el aumento de la prevalencia de la nefritis, la protección de *KIAA1542* frente a la manifestación del desorden inmunológico, la asociación de *UHRF1BP1* con el aumento de la prevalencia del desorden inmunológico y la protección de *ITGAM* frente a la artritis<sup>174</sup>.

Sin embargo, todos estos resultados muestran evidencias de asociación débiles y muy poco reproducibles. Esto puede ser debido a que al realizar este tipo de estudios se pierde poder estadístico al distribuir a los pacientes en base a sus fenotipos clínicos. Otra posible causa es que los *loci* estudiados provienen normalmente de su descubrimiento mediante GWAS donde los pacientes utilizados presentan una amplia heterogeneidad clínica lo que imposibilita la búsqueda de asociaciones con subfenotipos clínicos más claras. Por lo tanto, es necesaria la confirmación por medio de más estudios utilizando amplias colecciones con la información clínica detallada para afianzar estas asociaciones con los subfenotipos del LES.

**HIPÓTESIS.**

---



1. El LES es una enfermedad muy heterogénea tanto clínica como genéticamente. Estudios previos han observado que el riesgo a padecer diferentes manifestaciones clínicas de la enfermedad varía no sólo según el grupo étnico sino que también dentro del mismo. Al mismo tiempo, se ha descrito un fenómeno de subestructura genética dentro de la población europea que se puede definir mediante el estudio de marcadores informativos de ancestralidad (AIMs). Por lo tanto, nos planteamos la presencia de diferencias en la prevalencia de las manifestaciones clínicas del LES resultado de la subestructura europea.
2. Se han descrito más de 30 *loci* asociados con la susceptibilidad al LES. Varios de estos *loci* presentan resultados discordantes dependiendo no sólo de la etnia estudiada sino que también dentro de la misma. Por lo tanto, nos planteamos la presencia de diferencias en el tamaño de los efectos de múltiples *loci* asociados al LES resultado de la subestructura europea.
3. Varios polimorfismos presentes en el gen *IRF5* y regiones adyacentes regulan su expresión. Los estudios de expresión con *microarrays* realizados en colecciones de líneas linfoblastoides, aportan datos de expresión a nivel de genoma completo. Mediante el análisis de estos datos de expresión en relación a los polimorfismos de *IRF5*, podremos identificar aquellos polimorfismos que influyen la expresión del gen.
4. Debido a la falta de repercusión funcional aparente del SNP protector rs729302, nos planteamos la hipótesis de que este SNP funcionaría como *tag* SNP de otro polimorfismo desconocido situado en la región promotora de *IRF5*.
5. En relación a la existencia de nuevos polimorfismos funcionales del gen *IRF5* y región promotora, nos planteamos estudiar su papel en la explicación de la susceptibilidad al LES.

## **OBJETIVOS.**

---

1. Estudiar la influencia de la subestructura europea mediante el análisis de 6 AIMs en la prevalencia de los 11 criterios clínicos de clasificación del LES establecidos por la ACR además de en la edad de inicio de la enfermedad.
2. Analizar las diferencias en el tamaño del efecto de los alelos de riesgo de 25 *loci* asociados con LES en un total de 1742 pacientes y 2101 controles recopilados en 10 países europeos. Concretamente, estudiaremos las diferencias encontradas entre las colecciones del centro de Europa y las colecciones del sur de Europa.
3. Analizar la relación entre los datos de expresión de *IRF5* en estudios de *microarrays* realizados en líneas linfoblastoides y los polimorfismos del gen, para identificar los factores reguladores en *cis*.
4. Explorar la región promotora de *IRF5* en busca de polimorfismos funcionales que puedan explicar el efecto protector del SNP rs729302.
5. Determinar si los nuevos polimorfismos funcionales están asociados con la susceptibilidad al LES en una gran cohorte independiente de muestras europeas mediante un estudio de asociación caso-control.

## **PROCEDIMIENTO EXPERIMENTAL Y RESULTADOS.**

---

## PROCEDIMIENTO EXPERIMENTAL Y RESULTADOS.

### Esquema general:

#### 1. Heterogeneidad del LES.

##### 1.1. Heterogeneidad clínica del LES:

- Publicación 1: **Alonso-Perez E**, Suarez-Gestal M, Calaza M, Witte T, Papasteriades C, Marchini M, Migliaresi S, Kovacs A, Ordi-Ros J, Bijl M, Santos MJ, Ruzickova S, Pullmann R, Carreira P, Skopouli FN, D'Alfonso S, Sebastiani GD, Suarez A, Blanco FJ, Gomez-Reino JJ, Gonzalez A; European Consortium of SLE DNA Collections. *Association of systemic lupus erythematosus clinical features with European population genetic substructure*. **PLoS One**. 2011;6(12):e29033.

##### 1.2. Heterogeneidad genética del LES:

- Publicación 2: **Alonso-Perez E**, Suarez-Gestal M, Calaza M, Sebastiani GD, Pullmann R, Papasteriades C, Kovacs A, Skopouli FN, Bijl M, Suarez A, Marchini M, Migliaresi S, Carreira P, Ordi-Ros J, Witte T, Ruzickova S, Santos MJ, Barizzzone N, Blanco FJ, Lauwerys BR, Gomez-Reino JJ, Gonzalez A, Ecsledac EC. *Bias in effect size of Systemic Lupus Erythematosus susceptibility loci across Europe: a case control study*. **Arthritis Res Ther**. 2012 Apr 27;14(2):R94. [Epub ahead of print].

## 2. Estudio del gen *IRF5* en relación al LES.

### 2.1. Estudio de la variabilidad genética de *IRF5* en relación al LES:

- Publicación 3: Ferreiro-Neira I, Calaza M, **Alonso-Perez E**, Marchini M, Scorza R, Sebastiani GD, Blanco FJ, Rego I, Pullmann R Jr, Pullmann R, Kallenberg CG, Bijl M, Skopouli FN, Mavromati M, Migliaresi S, Barizzzone N, Ruzickova S, Dostal C, Schmidt RE, Witte T, Papasteriades C, Kappou-Rigatou I, Endreffy E, Kovacs A, Ordi-Ros J, Balada E, Carreira P, Gomez-Reino JJ, Gonzalez A. *Opposed independent effects and epistasis in the complex association of IRF5 to SLE*. **Genes Immun.** 2007 Jul;8(5):429-38.

### 2.2. Estudio de la regulación de la expresión génica de *IRF5* mediante datos de expresión procedentes de estudios con *microarrays*:

- Publicación 4: **Alonso-Perez E**, Suarez-Gestal M, Calaza M, Kwan T, Majewski J, Gomez-Reino JJ, Gonzalez A. *Cis-regulation of IRF5 expression is unable to fully account for systemic lupus erythematosus association: analysis of multiple experiments with lymphoblastoid cell lines*. **Arthritis Res Ther.** 2011 May 31;13(3):R80.

### 2.3. Mapa de LD de *IRF5*: búsqueda de polimorfismos funcionales:

2.3.1 Material y Métodos.

2.3.2 Resultados.

## **Heterogeneidad clínica del LES.**

# Association of Systemic Lupus Erythematosus Clinical Features with European Population Genetic Substructure

Elisa Alonso-Perez<sup>1\*</sup>, Marian Suarez-Gestal<sup>1\*</sup>, Manuel Calaza<sup>1</sup>, Torsten Witte<sup>2</sup>, Chryssa Papasteriades<sup>3</sup>, Maurizio Marchini<sup>4</sup>, Sergio Migliaresi<sup>5</sup>, Attila Kovacs<sup>6</sup>, Josep Ordi-Ros<sup>7</sup>, Marc Bijl<sup>8</sup>, Maria Jose Santos<sup>9</sup>, Sarka Ruzickova<sup>10</sup>, Rudolf Pullmann<sup>11</sup>, Patricia Carreira<sup>12</sup>, Fotini N. Skopouli<sup>13</sup>, Sandra D'Alfonso<sup>14</sup>, Gian Domenico Sebastiani<sup>15</sup>, Ana Suarez<sup>16</sup>, Francisco J. Blanco<sup>17</sup>, Juan J. Gomez-Reino<sup>1,18</sup>, Antonio Gonzalez<sup>1\*</sup>, and for the European Consortium of SLE DNA Collections

**1** Laboratorio de Investigacion, Instituto de Investigacion Sanitaria, Universitario de Santiago Hospital Clinico Santiago de Compostela, Santiago de Compostela, Spain, **2** Division of Clinical Immunology and Rheumatology, Department of Internal Medicine, Hannover Medical School, Hannover, Germany, **3** Department of Histocompatibility and Immunology, Evangelismos Hospital, Athens, Greece, **4** Clinical Immunology, Fondazione Istituto di Ricovero e Cura Carattere Scientifico (IRCCS) Ospedale Maggiore Policlinico, University of Milan, Milan, Italy, **5** Rheumatology Unit, Second University of Naples, Naples, Italy, **6** Department of Rheumatology, Hospital of Hungarian Railways, Szolnok, Hungary, **7** Internal Medicine Research Laboratory in Autoimmune Diseases, Hospital Vall d'Hebron, Barcelona, Spain, **8** Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, Groningen, The Netherlands, **9** Rheumatology Department, Hospital Garcia de Orta and Rheumatology Research Unit, Instituto de Medicina Molecular, Lisboa, Portugal, **10** Institute of Biotechnology, Academy of Sciences of the Czech Republic, Prague, Czech Republic, **11** Institute of Clinical Biochemistry, Martin Faculty Hospital, Jessenius Medical Faculty, Martin, Slovakia, **12** Rheumatology Department, Hospital 12 de Octubre, Madrid, Spain, **13** Pathophysiology Department, Athens University Medical School, Athens, Greece, **14** Department Medical Sciences and Interdisciplinary Research Center of Autoimmune Disease (IRCAD), Eastern Piedmont University, Novara, Italy, **15** Ospedale S. Camillo - Forlanini, Unità Operativa Complessa di Reumatologia, Roma, Italy, **16** Department of Functional Biology, University of Oviedo, Oviedo, Spain, **17** Laboratorio de Investigación Osteoarticular y del Envejecimiento, Servicio de Reumatología, Complejo Hospitalario Universitario A Coruña, A Coruña, Spain, **18** Department of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

## Abstract

Systemic Lupus Erythematosus (SLE) is an autoimmune disease with a very varied spectrum of clinical manifestations that could be partly determined by genetic factors. We aimed to determine the relationship between prevalence of 11 clinical features and age of disease onset with European population genetic substructure. Data from 1413 patients of European ancestry recruited in nine countries was tested for association with genotypes of top ancestry informative markers. This analysis was done with logistic regression between phenotypes and genotypes or principal components extracted from them. We used a genetic additive model and adjusted for gender and disease duration. Three clinical features showed association with ancestry informative markers: autoantibody production defined as immunologic disorder ( $P=6.8\times 10^{-4}$ ), oral ulcers ( $P=6.9\times 10^{-4}$ ) and photosensitivity ( $P=0.002$ ). Immunologic disorder was associated with genotypes more common in Southern European ancestries, whereas the opposite trend was observed for photosensitivity. Oral ulcers were specifically more common in patients of Spanish and Portuguese self-reported ancestry. These results should be taken into account in future research and suggest new hypotheses and possible underlying mechanisms to be investigated. A first hypothesis linking photosensitivity with variation in skin pigmentation is suggested.

**Citation:** Alonso-Perez E, Suarez-Gestal M, Calaza M, Witte T, Papasteriades C, et al. (2011) Association of Systemic Lupus Erythematosus Clinical Features with European Population Genetic Substructure. PLoS ONE 6(12): e29033. doi:10.1371/journal.pone.0029033

**Editor:** Pierre Bobé, Institut Jacques Monod, France

**Received:** September 8, 2011; **Accepted:** November 19, 2011; **Published:** December 14, 2011

**Copyright:** © 2011 Alonso-Perez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Instituto de Salud Carlos III (grants [06/0620, 08/0744]), by the RETICS Program (RD08/0075 (RIER)) which is partially financed by the European Regional Development Fund program of the European Union, and by grants from the Xunta de Galicia and by (DFG WI 1031/6-1) awarded to TW. EA-P is the recipient of a pre-doctoral bursary of the Instituto de Salud Carlos III. MS-G is the recipient of a FPU pre-doctoral bursary of the Spanish Ministry of Education. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: antonio.gonzalez.martinez-pedrayo@sergas.es

These authors contributed equally to this work.

## Introduction

Systemic Lupus Erythematosus is an autoimmune disease with a very varied spectrum of clinical manifestations [1]. It can affect multiple tissues and organs including kidneys, joints, skin, pleura and pericardium, diverse blood cells and the nervous system. It is also associated with a large variety of auto-antibodies and abnormalities of the immune system. These features are not present in all patients or at all times in the same patient. The disease course alternates flares and periods of remission and

clinical presentation can be different in subsequent flares from the observed previously in the same patient. This clinical heterogeneity poses many challenges to clinical diagnosis, treatment and research. Unfortunately, our understanding of its causes is still very incomplete, although it seems that genetic, environmental and socioeconomic factors have a role.

Recent Genome Wide Association studies (GWAS) have provided a list of more than 30 confirmed SLE susceptibility loci [2]. Some of them have been associated with particular SLE clinical features, but they are far to explain its clinical



heterogeneity [3]. Other studies have pointed to a broad effect of genetics in the form of the specific genetic background of human subpopulations. There has been knowledge of differences in SLE phenotype between continental ethnic groups for decades [1,4], but only research in recent years has been able to confirm the importance of genetic background by discriminating between genetics and socioeconomic or environmental factors [5,6,7]. The demonstration of an effect of genetic background in SLE phenotype provides the foundation for exploring the possibility that substructure within an ethnic group could influence also the disease clinical presentation. Recent work in about 1900 European-American SLE patients seems to support this hypothesis by showing correlation between the prevalence of some clinical features and ancestry informative markers (AIMs) [8,9]. These markers are SNPs that had shown in previous studies large differences in allele frequency between Europeans from different ancestries [10,11,12,13]. The finding of these correlations between clinical features and European substructure is very important to discriminate between the different factors influencing SLE heterogeneity and it is possible it could increase our power to identify etiological relationship for the different SLE phenotypes.

Our aim has been to explore the influence of European population substructure in the SLE phenotype of about 1400 European SLE patients from 9 countries. Three of the 12 clinical features analyzed, production of autoantibodies, oral ulcers and photosensitivity, were associated with informative European AIMs confirming the likely effect of variation in genetic background within the European ethnicity.

## Materials and Methods

### Ethic statement

All patients gave their written informed consent to participate and sample collection and study was approved by the relevant ethics committees at each of the recruiting centres. The project was approved by the Comité de Investigación Clínica de Galicia (Spain).

### Patient data

Samples from 1413 European SLE patients recruited at 16 centres from nine different countries were collected as described [14]. Patients were questioned about their ancestry and only patients with uniform ancestry from the country of origin were included. Data retrieved from each patient included the 11 SLE American College of Rheumatology (ACR) classification criteria [15], age of disease onset, disease duration and gender (**Table 1**). Immunologic disorder was defined as for these criteria and included mainly anti-dsDNA or anti-Sm positive patients. Antinuclear antibodies (ANA) were not included in the analysis because they were almost uniformly present in all patients. Each recruiting centre contributed a mean of 88.3 SLE patients with range from 58 to 128 patients (**Figure 1**).

### Genotyping of AIMs

Six AIMs were determined in the SLE patients: rs6730157, rs382259, rs4988235, rs12203592, rs354690 and rs12913832. The three first are the most informative AIMs in differentiating Northern from Southern European subpopulations identified in a study analyzing 300 000 SNPs in 4000 European subjects [11]. Results from rs4988235 were not used for analysis because it was largely redundant with rs6730157 in our samples ( $r^2 = 0.87$ ). rs12203592, rs354690 are the two AIMs more informative for East-West place of origin inside Europe according to the same study [11]. rs12913832 is a SNP associated with large differences in frequency across

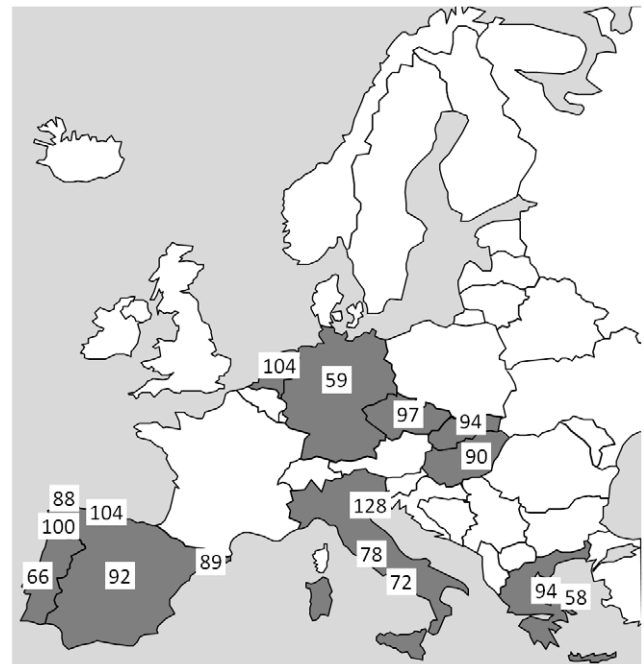
**Table 1.** Clinical characteristics of the patients with SLE.

Characteristic <sup>a</sup>	% (95% C.I.)	mean $\pm$ S.D.
Women	89.5 (87.9–91.1)	
Age of onset		31.1 $\pm$ 13.1
Disease duration		11.9 $\pm$ 8.3
Malar rash	55.7 (53.0–58.4)	
Discoid rash	17.8 (15.8–19.8)	
Photosensitivity	52.4 (49.7–55.1)	
Oral ulcers	28.0 (25.6–30.4)	
Arthritis	80.3 (78.2–82.4)	
Serositis	35.5 (33.0–38.0)	
Renal disorder	40.5 (37.9–43.1)	
Neurologic disorder	13.6 (11.8–15.5)	
Hematological disorder	71.3 (68.9–73.7)	
Immunological disorder <sup>b</sup>	78.7 (76.5–80.9)	
ANA	91.4 (89.8–93.0)	

<sup>a</sup>Data from >98% of the patients for all characteristics except for the following: malar rash, neurologic disorder, hematologic disorders and age of disease onset with data from >92% of the patients; disease duration and ANAs that were not available from two recruiting centres (available in >80%).

<sup>b</sup>Defined as for the SLE ACR classification criteria [15] including abnormal anti native DNA, anti-Sm antibodies, LE cells or false positive serologic syphilis test. doi:10.1371/journal.pone.0029033.t001

Europe and unrelated with the previous [10]. These 5 SNPs were amplified in a single PCR reaction done with the KAPA2G fast HotStart (Kapa Biosystems, Woburn MA, USA) on a final volume of 10  $\mu$ l (20 ng genomic DNA), using 3 mM MgCl<sub>2</sub> and 0.2  $\mu$ M of each primer. Products were purified by Exo-SAP digestion with Exonuclease I (Epicentre, Madison, WI) and Shrimp Alkaline



**Figure 1.** Collections of SLE patients with number of patients available for analysis.

doi:10.1371/journal.pone.0029033.g001

Phosphatase (GE Healthcare, Barcelona, Spain). Subsequently, single-base extension reactions with the SNaPshot Multiplex kit (Applied Biosystems, Foster City, CA) were done. Samples were analyzed in an AbiPrism 3130xl Genetic Analyzer (Applied Biosystems). Genotyping call rate success was 99.7%. Sequences of primers and probes are available from the authors upon request.

### Statistical analysis

We have computed the allelic frequencies of the AIMs per each of the 16 recruiting centres to assess their variability and whether they follow the previously reported trends along Europe. Concordance of these genotypes with Hardy-Weinberg equilibrium (HWE) was also assessed by each of the centres given that one of the causes of deviation is population stratification and the AIMs have population specific frequencies. The *P* value for claiming deviation from HWE was set at 0.01, a conservative threshold taking into account the number of centres and AIMs. Factor analysis via principal component (PC) extraction was applied to the AIM genotypes to reduce dimensionality. Association of each of the ACR classification criteria with each of the AIMs and with the retained principal components was analyzed by logistic regression. Genotypes were coded according to an additive model (0, 1 and 2, for the common homozygote, the heterozygote and the rare homozygote genotypes, respectively). The results that are presented included gender and disease duration as covariates. The odds ratio by each allele (O.R.) and their 95% confidence intervals (C.I.) are also given. Analysis of association of age of disease onset with the AIMs genotypes or with the PC was done with multiple linear regression. Genotypes were coded in a similar way, and gender was included as a covariate in these analyses. Analyses were also conducted with inclusion of the recruitment centre as covariate and without any covariates and if results changed interpretation, this circumstance was reported. All statistical analyses were done with Statistica 7.0 (StatSoft, Inc., Tulsa, OA). A significance threshold of 0.008 was applied according to a Bonferroni correction for the six AIMs analyzed.

## Results

### AIM genotypes and population substructure

None of the five AIMs was significantly deviated from HWE in any of the sample collections ( $P > 0.01$ ). They showed a large variation between patients with different self-reported ancestry within Europe (**Table 2**). The most extreme difference was observed for rs6730157 that showed an A allelic frequency of

15.5% in Greek patients, and of 73.6% in Dutch patients. The most restricted range of frequencies was observed for rs354690 (from 37.1% to 45.9% frequency of the T allele). Three of the AIMs, rs6730157, rs12913832 and rs382259, showed a clear differentiation between patients from Southern European countries (Portugal, Spain, Italy and Greece) and those from Central Europe (The Netherlands, Germany, Czech Republic, Slovakia and Hungary). This is in agreement with major axis of known European population substructure [11,12,13].

We applied factor analysis to the genotypes of the five AIMs and two PC explaining 46.5% of the variance were retained. Rs6730157 was the main contributor to PC1, whereas rs12203592 was the main contributor to PC2. These PC showed significant correlations with the geographical coordinates of the patient's ancestries (**Table 3**): PC1 with the latitude ( $r = -0.47$ ) but not with the longitude; and PC2 correlated with both the latitude ( $r = -0.13$ ) and the longitude ( $r = 0.10$ ). Some of the individual AIMs showed a stronger correlation with these coordinates than others and the correlations were particularly strong with latitude (Table 3).

### Association of SLE clinical features with population substructure

Once we had confirmed that the five AIMs were informative for European population substructure in our patients, we used them to look for evidence of its effects in the phenotype of SLE. Three of the ACR classification criteria showed association with some AIMs. The presence of the immunological disorder criterion, which consists in production of a variety of specific autoantibodies (mainly anti-dsDNA or anti-Sm antibodies), was associated with rs382259 ( $P = 6.8 \times 10^{-4}$ ; O.R. = 0.70, 95% C.I. = 0.57–0.86). Increased prevalence of immunological disorder was associated with the CC and TC genotypes of rs382259 (**Figure 2**). Association was also found between immunological disorder and PC1 ( $P = 0.004$ ; O.R. = 0.82, 95% C.I. = 0.72–0.94).

Oral ulcers were also associated with two AIMs: rs12913832 ( $P = 6.9 \times 10^{-4}$ ; O.R. = 0.73, 95% C.I. = 0.61–0.87) and rs382259 ( $P = 9.3 \times 10^{-4}$ ; O.R. = 0.72, 95% C.I. = 0.59–0.87) in single AIM analyses. Association with these two AIMs persisted in multivariate analysis that included the five AIMs ( $P = 2.0 \times 10^{-4}$  and  $P = 1.7 \times 10^{-3}$ , respectively) indicating that each of the two AIMs have an independent contribution to the association. Higher prevalence of oral ulcers was associated with the AA and GA genotypes of rs12913832, and with the TT but not with the TC genotypes of rs382259 (**Figure 2**). No association with the PCs

**Table 2.** Allele frequency of the ancestry informative markers (AIMs) by country of self-reported ancestry.

Country <sup>a</sup>	rs6730157 A	rs12913832 G	rs382259 T	rs12203592 T	rs354690 T
Greece	0.155	0.367	0.439	0.037	0.452
Italy	0.160	0.417	0.485	0.113	0.435
Portugal	0.386	0.258	0.689	0.075	0.409
Slovakia	0.403	0.699	0.688	0.102	0.441
Hungary	0.426	0.670	0.646	0.084	0.371
Spain	0.430	0.329	0.678	0.145	0.407
Czech R.	0.479	0.792	0.768	0.149	0.459
Germany	0.578	0.767	0.698	0.121	0.397
Netherlands	0.736	0.808	0.798	0.067	0.4237

<sup>a</sup>Countries are ordered according to allele frequencies of rs6730157. The SNPs are from most variable to less variable, left to right.

doi:10.1371/journal.pone.0029033.t002

**Table 3.** Correlation between the two first principal components (PC) obtained from the AIM genotypes and between each of the AIMS with the geographic coordinates of the reported ancestries of the SLE patients.

PC or AIM	Latitude		Longitude	
	r	P	r	P
PC1	0.47	$<10^{-6}$	0.03	ns
PC2	-0.13	$10^{-6}$	0.10	$1.4 \times 10^{-4}$
rs6730157	0.71	0.002	-0.40	ns
rs382259	-0.62	0.010	0.50	0.048
rs12913832	0.91	$10^{-6}$	0.42	ns
rs12203592	0.20	ns	-0.58	0.019
rs354690	-0.07	ns	0.41	ns

doi:10.1371/journal.pone.0029033.t003

was observed. Analysis including the recruiting centers as covariates showed that association with the two AIMS was completely dependent on this factor. This was due to the higher frequency of oral ulcers in patients from Spain and Portugal according to their self-reported ancestry than in patients from other origins (47.1% versus 18.5%;  $P < 10^{-6}$ ).

Finally, photosensitivity was associated with rs12203592 ( $P = 0.0021$ ; O.R. = 1.47, 95% C.I. = 1.15–1.88). This association was independent of conditional analysis with the other four AIMS ( $P = 0.001$ ). Photosensitivity was more common in patients that were TT or CT for rs12203592 or (Figure 2).

The remaining SLE ACR criteria and the age of disease onset were not associated with any AIM or PC.

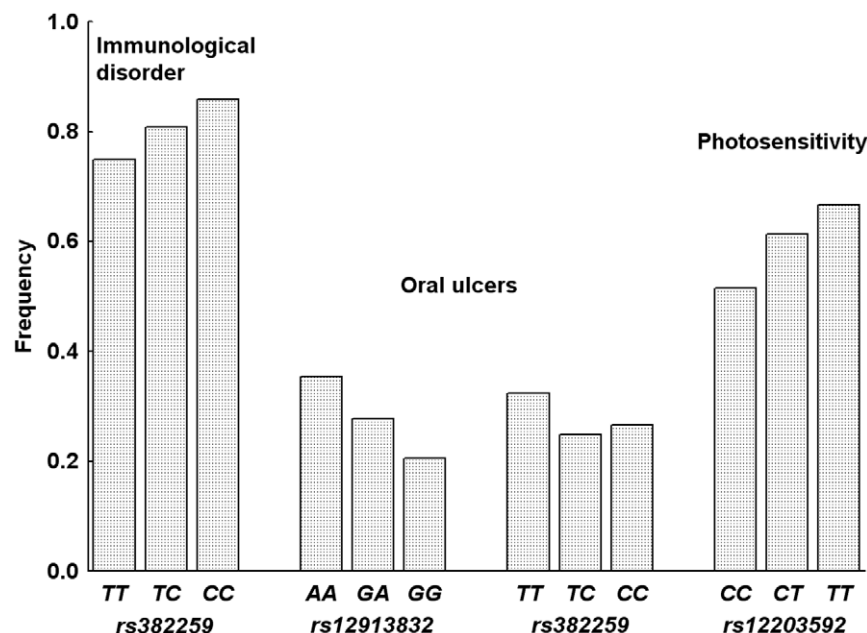
## Discussion

Our results have confirmed a significant effect of European population substructure on the SLE phenotype. The substructure

associated phenotypes, immunological disorder, oral ulcers and photosensitivity, have already been identified in the only other group of SLE patients in which this possibility has been tested [8,9]. The consistency of results adds credibility to the findings. However, we have not replicated association of other clinical features from the previous study and the effect on oral ulcers was not fully convincing, either in our study or in the previous one [8,9].

The AIMS we have used were able to show European population substructure. Their variation was mainly marked by the North-South differentiation that has been found in previous studies [10,11,12,13]. This is remarkable because our analysis included only a fraction of the many European populations used in these studies. Of potential relevance is the lack in our study of Scandinavian subjects or of subjects from the British Islands or from Russia that had been included in the studies for the discovery of the AIMS [11,12,13]. These populations that correspond to geographical extremes to the North, West and East of the European population could enlarge the range of AIM frequencies and improve correlation of these frequencies with geographical coordinates.

A limitation of the AIMS we have used is that they are not enough for classification of individual subjects. However, we think they were sufficient to detect a large fraction of the SLE clinical features associated with European population substructure. This conclusion is based in two pieces of evidence. The first is that the three findings of our study were associated with more than one AIM indicating a certain level of redundancy in spite of the low correlation between the AIM genotypes (mean pairwise  $r^2 = 0.007$ ). The second is that only the very top AIMS contributing to the first two PCs for population substructure in Europeans, according to a study with 300 000 SNPs [11], were informative in our study: the third AIM contributing to PC1, rs4988235, showed a high correlation with the first AIM, rs6730157 ( $r^2 = 0.87$ ), and the same pattern of associations (not shown); and rs354690 that was the second AIM for PC2 did not show association with any of the clinical features.



**Figure 2.** Frequency of the SLE clinical features associated with AIM genotypes. The abscise axis indicates the AIM genotype and in the ordinate axis are the frequencies of the indicated clinical manifestation.  
doi:10.1371/journal.pone.0029033.g002

Association of the production of the SLE specific autoantibodies that are included in the immunologic disorder criterion (in most cases, antibodies to dsDNA or the nuclear Sm antigen) with rs382259 and with PC1 indicates a role of the patient's genetic background. It is worth to mention that immunological disorder was also associated with the two first PCs in the study done in European American SLE patients [9]. This association was interpreted as meaning that a Southern and Western European ancestry predisposes to autoantibody production. Our results can be interpreted also as meaning that a Southern European ancestry predisposes to this phenotype, but no clear differentiation in the West-East axis was observed.

The effect of genetic background in the prevalence of autoantibodies had already been shown in relation with the continental ethnic groups. For example, there are reports showing an increased prevalence of anti-Sm and anti-RNP antibodies in African patients relative to Europeans [16,17,18,19]; other differences concern clusters of autoantibodies that include anti-Sm or anti-dsDNA [20], or other specific SLE autoantibodies like anti-P [21] or anti-RNA helicase A [18,22]. But, there is not any report of a general higher prevalence of SLE autoantibodies in patients of a specific ethnic group. On the contrary, some autoantibodies have been found at low prevalence while others show high prevalence in the same ethnicity [18,20,22]. Therefore, it seems likely that genetic background is influencing specific responses more than the general abnormalities leading to antibody mediated autoimmunity. This could be the case for HLA alleles whose frequency is highly variable between populations and that affect prevalence of anti-P [21], anti-cardiolipin and anti-beta2GPI antibodies in different ethnicities [23]. It seems likely that differences within Europeans affect autoantibody production in SLE patients in the same way.

The association of oral ulcers with European population substructure is more open to question. Although we have observed a clear association with two AIMs, rs12913832 and rs382259, it disappeared after adjusting by center of recruitment. This result invites to caution but does not invalidate interpretation because centers of recruitment were strongly linked with the patient's reported ancestries. In fact, classifying the patients by their reported ancestries showed a clear excess of oral ulcers in patients from Spain and Portugal. This increased prevalence is reflected by with the AIMs association with a higher frequency of the associated genotypes in the South-Western European ancestries. Results from the study on European American patients showed an increase of oral ulcers also in patients with Southern European ancestry, but again it did not persist after adjusting for covariates [9]. Given these results, we cannot conclude at present. We cannot distinguish a genuine difference in the phenotype of patients with a South-Western European ancestry from confounding factors associated with recruiting hospitals. In addition, there has not been a wide interest in the analysis of variation in prevalence of oral ulcers between SLE patients from different ethnicities. Large differences between SLE patient series from different ethnicities have been reported [4], but as they were not obtained in comparative studies and have not been replicated, it is unclear whether these differences could be attributable to genetic background. Therefore, we lack evidence of reproducible trends and of possible etiologic factors that could help us to interpret the current results. However, there are data from large collections of European SLE patients that support the difference we have found: two collections of Spanish SLE patients showed a prevalence of oral ulcers of 46.4% and 54.3% (of 462 and 490 patients, respectively) [24], and a study of 544 Portuguese SLE patients reported a prevalence of 45% [25]; whereas the Euro-Lupus study

showed a prevalence of 12.5% (of 1000 patients from all over Europe) [26].

The third SLE clinical feature we have found associated with European population substructure is photosensitivity. It was associated with the genotypes of two AIMs. rs12913832 direction was characteristic of Northern Europeans. An excess of photosensitivity in SLE patients with Northern European ancestry was also found in the previous study of European-Americans [8,9]. This effect of population substructure is the most amenable to interpretation because it could be related with lighter skin pigmentation, which has a key role in sensitivity to sunburns, melanoma and other UV-related cancers and which is much more common in Northern Europeans than in Southern Europeans. This interpretation is in agreement with the lower prevalence of SLE photosensitivity among African American patients with SLE [19], or black patients from South Africa [27] or Jamaica [28]. However, rs12203592 was also associated with photosensitivity without any discernible geographical frequency distribution, and no studies of skin pigmentation or phototype in relation with SLE photosensitivity have been done in Europeans. Therefore, this hypothesis requires specific testing. It could be done directly by comparing prevalence of photosensitivity in function of skin color and sun exposure, but also by looking for association between the wide array of loci already known to determine skin pigmentation, which show wide differences in frequency among Europeans [10,29], and SLE photosensitivity.

Other SLE clinical features showed association with European population substructure in the Richman et al. study [9], but not in our study and, therefore, they remain unconfirmed. These include discoid rash, renal disorder, serositis, neurological disorder and malar rash. Some of them were weakly associated in the Richman study, like malar rash and neurological disorder, but the others showed *P* values below 0.01. These differences could be due to lack of power of our study, false positive findings in the Richman study or to differences between SLE patients and study design of the two reports. For example, our study included a high fraction of subject with self-reported ancestry from Southern European countries followed by Central European countries, whereas the European American patients were very markedly of Northern and Western European ancestry [8].

As has been commented already, the clinical features associated with European populations substructure could be due to their relation with SLE loci that show a frequency gradient within Europe. Apart from the HLA alleles [30], we know already of other differences in SLE associated loci between Europeans from different ancestries. The SLE risk allele of PTPN22 R620W shows a higher frequency in subjects from the North and West of Europe than in those from the South [31]. Our Consortium has also shown that a difference in frequencies, although more complex, is present for PD1.3, a SLE risk polymorphism the PDCD1 locus [32,33]. The A allele is more common in SLE patients than in controls from the North Center of Europe, similar in patients and controls from the Southeast, and more common in controls than in SLE patients from the Southwest of Europe [33]. Notably, the gradient in frequency has been observed only in controls, not in patients. It is possible than these and similar variations in risk allele frequencies explain the associations between SLE clinical features and European population substructure.

One of the limitations of our study is that it does not include a representative sample of SLE patients from all the countries with European ancestry. It will be also desirable to include in the analysis other important characteristics of SLE besides the ACR classification criteria like additional clinical manifestations or any of the damage or disease activity indexes. It will be also beneficial

to count with more detailed information about autoantibodies or severity of the different clinical features. A concerted effort will be necessary to obtain this type of high quality data from a large number of patients of multiple European ancestries.

We have not corrected the *P* values for number clinical features analyzed. This was motivated by the difficulty in defining an appropriate level of correction. Two issues make this difficult: the correlation between clinical features that will lead to overcorrection if a Bonferroni approach is used; and the contentious issue of whether association of each clinical feature should be considered as an independent hypothesis or as a unique hypothesis [34,35].

In summary, our study reinforces the evidence of a significant but modest degree of variation in the SLE phenotype in relation with European population substructure. The differences we have found will help to understand the underlying mechanisms as clues of possible association with particular loci and skin pigmentation are already suggested. Elucidation of these mechanisms will advance our ability to cope with SLE clinical heterogeneity. In addition, this finding is important for the design of SLE clinical projects including patients of European ancestry that until now have been taken as a unit of study or comparison. Accounting for European substructure will be even more important for research aiming to define the relationships between SLE phenotype and genotype. Specifically, our study contributes also to consolidate the association between Southern European ancestry and more prevalence of autoantibody production and less photosensitivity in SLE patients and suggests the possibility of an increased frequency of oral ulcers in patients with South-Western European ancestry.

## Acknowledgments

Other contributors to the European Consortium of SLE DNA Collections: Myriam Liz, Laboratorio de Investigacion 10 and Rheumatology Unit,

Instituto de Investigacion Sanitaria – Hospital Clinico Universitario de Santiago, Santiago de Compostela, Spain; Reinhold E Schmidt, Division of Clinical Immunology, Department of Internal Medicine of the Hannover Medical School, Hannover, Germany; Iris Kappou-Rigatou, Department of Histocompatibility and Immunology, Evangelismos Hospital, Athens, Greece; Raffaella Scorza, Clinical Immunology, University of Milan and Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy; Emöke Endereff, Department of Paediatrics, Albert Szent-Györgyi Clinical Centre, University of Szeged, Szeged, Hungary; Eva Balada, Internal Medicine, Research Laboratory in Autoimmune Diseases Hospital Vall d'Hebron, Barcelona, Spain; Cees G Kallenberg, Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, Groningen, The Netherlands; Filipe Vinagre, Rheumatology Department, Hospital Garcia de Orta, Almada (Portugal) and Rheumatology Research Unit, Instituto Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa (Portugal); Căbitor Dostal, Molecular Biology and Immunogenetics Dept., Institute of Rheumatology, Prague, Czech Republic; Rudolf Pullmann Jr, Gerontology Research Center, National Institute on Aging, Nathan Shock Drive, Baltimore, Maryland, USA; Maria Mavromati, Pathophysiology Department, Athens University Medical School, Athens, Greece; Nadia Barizzzone, Dept. Medical Sciences and IRCAD, Eastern Piedmont University, Novara, Italy; Carmen Gutierrez, Department of Functional Biology, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo, Spain; Ignacio Rego, Laboratorio de Investigación Osteoarticular y del Envejecimiento, Servicio de Reumatología, CH Universitario A Coruña, A Coruña, Spain.

## Author Contributions

Conceived and designed the experiments: EA-P MS-G MC JJG-R AG. Performed the experiments: EA-P MS-G TW CP MM SM AK JO-R MB MJS SR RP PC FNS SDA GDS AS FJB JJG-R. Analyzed the data: EA-P MS-G MC AG. Contributed reagents/materials/analysis tools: TW CP MM SM AK JO-R MB MJS SR RP PC FNS SDA GDS AS FJB JJG-R. Wrote the paper: EA-P MS-G MC TW CP MM SM AK JO-R MB MJS SR RP PC FNS SDA GDS AS FJB JJG-R AG.

## References

- Wallace DJ, Hahn B, Dubois EL (2007) Dubois' lupus erythematosus. Philadelphia: Lippincott Williams & Wilkin. xvii, 1414 p.
- Deng Y, Tsao BP (2010) Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat Rev Rheumatol* 6: 683–692.
- Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, et al. (2011) Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. *PLoS Genet* 7: e1001311.
- Tikly M, Navarra SV (2008) Lupus in the developing world—is it any different? *Best Pract Res Clin Rheumatol* 22: 643–655.
- Seldin MF, Qi L, Scherbarth HR, Tian C, Ransom M, et al. (2008) Amerindian ancestry in Argentina is associated with increased risk for systemic lupus erythematosus. *Genes Immun* 9: 389–393.
- Alarcon GS, Calvo-Alen J, McGwin G, Jr., Uribe AG, Toloza SM, et al. (2006) Systemic lupus erythematosus in a multiethnic cohort: LUMINA XXXV. Predictive factors of high disease activity over time. *Ann Rheum Dis* 65: 1168–1174.
- Alarcon GS, Bastian HM, Beasley TM, Roseman JM, Tan FK, et al. (2006) Systemic lupus erythematosus in a multi-ethnic cohort (LUMINA) XXXII: [corrected] contributions of admixture and socioeconomic status to renal involvement. *Lupus* 15: 26–31.
- Chung SA, Tian C, Taylor KE, Lee AT, Ortmann WA, et al. (2009) European population substructure is associated with mucocutaneous manifestations and autoantibody production in systemic lupus erythematosus. *Arthritis Rheum* 60: 2448–2456.
- Richman IB, Chung SA, Taylor KE, Kosoy R, Tian C, et al. (2010) European population substructure correlates with systemic lupus erythematosus endophenotypes in North Americans of European descent. *Genes Immun* 11: 515–521.
- Han J, Kraft P, Nan H, Guo Q, Chen C, et al. (2008) A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet* 4: e1000074.
- Tian C, Kosoy R, Nassir R, Lee A, Villoslada P, et al. (2009) European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol Med* 15: 371–383.
- Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, et al. (2006) European population substructure: clustering of northern and southern populations. *PLoS Genet* 2: e143.
- Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, et al. (2008) Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet* 4: e4.
- Suarez-Gestal M, Calaza M, Endreffy E, Pullmann R, Ordi-Ros J, et al. (2009) Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. *Arthritis Res Ther* 11: R69.
- Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 40: 1725.
- Tikly M, Burgin S, Mohanlal P, Bellingan A, George J (1996) Autoantibodies in black South Africans with systemic lupus erythematosus: spectrum and clinical associations. *Clin Rheumatol* 15: 261–265.
- Garcia CO, Molina JF, Gutierrez-Urena S, Scopelitis E, Wilson WA, et al. (1996) Autoantibody profile in African-American patients with lupus nephritis. *Lupus* 5: 602–605.
- Yamasaki Y, Narain S, Yoshida H, Hernandez L, Barker T, et al. (2007) Autoantibodies to RNA helicase A: a new serologic marker of early lupus. *Arthritis Rheum* 56: 596–604.
- Sestak AL, Nath SK, Kelly JA, Bruner GR, James JA, et al. (2008) Patients with familial and sporadic onset SLE have similar clinical profiles but vary profoundly by race. *Lupus* 17: 1004–1009.
- To CH, Petri M (2005) Is antibody clustering predictive of clinical subsets and damage in systemic lupus erythematosus? *Arthritis Rheum* 52: 4003–4010.
- Arnett FC, Reveille JD, Moutsopoulos HM, Georgescu L, Elkon KB (1996) Ribosomal P autoantibodies in systemic lupus erythematosus. Frequencies in different ethnic groups and clinical and immunogenetic associations. *Arthritis Rheum* 39: 1833–1839.
- Vazquez-Del Mercado M, Palafox-Sanchez CA, Munoz-Valle JF, Orozco-Barocio G, Oregon-Romero E, et al. (2010) High prevalence of autoantibodies to RNA helicase A in Mexican patients with systemic lupus erythematosus. *Arthritis Res Ther* 12: R6.
- Galeazzi M, Sebastiani GD, Tincani A, Piette JC, Allegri F, et al. (2000) HLA class II alleles associations of anticardiolipin and anti-beta2GPI antibodies in a large series of European patients with systemic lupus erythematosus. *Lupus* 9: 47–55.
- (2006) [Registrolasaf. Features of patients with systemic lupus erythematosus and antiphospholipid syndrome registered by rheumatologists and internists]. *Reumatol Clin* 2: 131–136.

25. Santos MJ, Capela S, Figueira R, Nero P, Matos AA, et al. (2007) [Characterization of a Portuguese population with systemic lupus erythematosus]. *Acta Reumatol Port* 32: 153–161.
26. Cervera R, Abarca-Costalago M, Abramovicz D, Allegri F, Annunziata P, et al. (2006) Systemic lupus erythematosus in Europe at the change of the millennium: lessons from the “Euro-Lupus Project”. *Autoimmun Rev* 5: 180–186.
27. Sutej PG, Gear AJ, Morrison RC, Tikly M, de Beer M, et al. (1989) Photosensitivity and anti-Ro (SS-A) antibodies in black patients with systemic lupus erythematosus (SLE). *Br J Rheumatol* 28: 321–324.
28. Smikle MF, Barton EN, Morgan OS, Deaculaer K (1996) Photosensitivity and antinuclear antibodies in black patients with systemic lupus erythematosus. *J Assoc Acad Minor Phys* 7: 53–55.
29. Nan H, Kraft P, Hunter DJ, Han J (2009) Genetic variants in pigmentation genes, pigmentary phenotypes, and risk of skin cancer in Caucasians. *Int J Cancer* 125: 909–917.
30. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR (2011) Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. *Nucleic Acids Res* 39: D913–919.
31. Gregersen PK, Lee HS, Batliwalla F, Begovich AB (2006) PTPN22: setting thresholds for autoimmunity. *Semin Immunol* 18: 214–223.
32. Ferreira-Vidal I, Gomez-Reino JJ, Barros F, Carracedo A, Carreira P, et al. (2004) Association of PDCD1 with susceptibility to systemic lupus erythematosus: evidence of population-specific effects. *Arthritis Rheum* 50: 2590–2597.
33. Ferreira-Vidal I, D’Alfonso S, Papasteriades C, Skopouli FN, Marchini M, et al. (2007) Bias in association studies of systemic lupus erythematosus susceptibility due to geographical variation in the frequency of a programmed cell death 1 polymorphism across Europe. *Genes Immun* 8: 138–146.
34. Bender R, Lange S (2001) Adjusting for multiple testing—when and how? *J Clin Epidemiol* 54: 343–349.
35. Feise RJ (2002) Do multiple outcome measures require p-value adjustment? *BMC Med Res Methodol* 2: 8.

## **Heterogeneidad genética del LES.**



RESEARCH ARTICLE

Open Access

# Bias in effect size of systemic lupus erythematosus susceptibility loci across Europe: a case-control study

Elisa Alonso-Perez<sup>1</sup>, Marian Suarez-Gestal<sup>1</sup>, Manuel Calaza<sup>1</sup>, Gian Domenico Sebastiani<sup>2</sup>, Rudolf Pullmann<sup>3</sup>, Chryssa Papasteriades<sup>4</sup>, Attila Kovacs<sup>5</sup>, Fotini N Skopouli<sup>6</sup>, Marc Bijl<sup>7</sup>, Ana Suarez<sup>8</sup>, Maurizio Marchini<sup>9</sup>, Sergio Migliaresi<sup>10</sup>, Patricia Carreira<sup>11</sup>, Josep Ordi-Ros<sup>12</sup>, Torsten Witte<sup>13</sup>, Sarka Ruzickova<sup>14</sup>, Maria Jose Santos<sup>15</sup>, Nadia Barizzone<sup>16</sup>, Francisco J Blanco<sup>17</sup>, Bernard R Lauwerys<sup>18</sup>, Juan J Gomez-Reino<sup>1,19</sup> and Antonio Gonzalez<sup>1\*</sup>, for the European Consortium of SLE DNA Collections

## Abstract

**Introduction:** We aimed to investigate whether the effect size of the systemic lupus erythematosus (SLE) risk alleles varies across European subpopulations.

**Methods:** European SLE patients ( $n = 1,742$ ) and ethnically matched healthy controls ( $n = 2,101$ ) were recruited at 17 centres from 10 different countries. Only individuals with self-reported ancestry from the country of origin were included. In addition, participants were genotyped for top ancestry informative markers and for 25 SLE associated SNPs. The results were used to compare effect sizes between the Central European and Southern European subgroups.

**Results:** Twenty of the 25 SNPs showed independent association with SLE. These SNPs showed a significant bias to larger effect sizes in the Southern subgroup, with 15/20 showing this trend ( $P = 0.019$ ) and a larger mean odds ratio of the 20 SNPs (1.46 vs. 1.34,  $P = 0.02$ ) as well as a larger difference in the number of risk alleles (2.06 vs. 1.63,  $P = 0.027$ ) between SLE patients and controls than for Central Europeans. This bias was reflected in a very significant difference in the cumulative genetic risk score (4.31 vs. 3.48,  $P = 1.8 \times 10^{-32}$ ). Effect size bias was accompanied by a lower number of SLE risk alleles in the Southern subjects, both patients and controls, the difference being more marked between the controls ( $P = 1.1 \times 10^{-8}$ ) than between the Southern and Central European patients ( $P = 0.016$ ). Seven of these SNPs showed significant allele frequency clines.

**Conclusion:** Our findings showed a bias to larger effect sizes of SLE loci in the Southern Europeans relative to the Central Europeans together with clines of SLE risk allele frequencies. These results indicate the need to study risk allele clines and the implications of the polygenic model of inheritance in SLE.

## Introduction

The systemic lupus erythematosus (SLE) genetic component has been partially elucidated thanks to large studies that have uncovered more than 30 loci reaching very convincing disease association [1-12]. These studies have shown that a large fraction of the SLE loci (such as

*STAT4*, *TNFSF4* or *BLK*) are shared in the different ethnic groups; however, other loci are not (such as *PTPN22*, which is exclusive of Europeans). These latter loci can be due to the absence or rarity of the polymorphism in one of the ethnic groups (as for *PTPN22*, which is absent in Asians), but other SLE loci show a similar frequency in discordant populations (as for *PXK* or *FCGR2A*). Possible explanations for these conflicting results have been envisaged, including differences in linkage between the causal polymorphism and the analysed SNPs, limitations of

\* Correspondence: antonio.gonzalez.martinez-pedrayo@sergas.es

<sup>1</sup>Laboratorio de Investigacion 10 and Rheumatology Unit, Instituto de Investigacion Sanitaria - Hospital Clinico Universitario de Santiago, Choupana s/n, Santiago de Compostela 15706, Spain

Full list of author information is available at the end of the article



study design, or differences in the interactions with other genetic loci or with environmental exposures [13].

Gene-gene interaction could be behind the observation that SLE loci show variability in effect sizes in function of the genetic background. For example, the Amerindian genetic background is associated with a higher effect size of several SLE loci in Hispanics [14]. We wondered whether genetic heterogeneity of this type could exist among European subjects. Support for this hypothesis is provided by the recent evidence of differences in SLE clinical features among Europeans [15-17] and by opposed results of SLE association with *PDCDI* [18,19]. Notably, these two observations showed a North-South axis of variations, which is the main axis of European population differentiation [20-23].

Our study included 25 top-associated SNPs in the better known SLE loci studied in 1,742 patients with SLE and in 2,101 controls from 17 collections recruited in 10 European countries, each of them with homogeneous local ancestry. The results showed a bias to larger effect sizes of the risk alleles in the Southern Europeans relative to the Central Europeans. We also found clines of risk allele frequencies.

## Materials and methods

### Patient data

We used DNA samples from European SLE patients and ethnically matched healthy controls recruited at 17 centres from 10 different countries (Table 1). Most of these samples have already been described [24,25]. Each recruiting centre was asked for about 100 patients with SLE according to the revised American College of Rheumatology classification criteria [26] and for about 100 controls, producing a total of 1,742 cases and 2,101 controls. Only individuals with self-reported ancestry from the country of origin were included. All participants gave their written informed consent to participate and the study was approved by the relevant ethics committees.

### Genotyping

DNA samples were amplified in a multiplex PCR with the KAPA2G fast HotStart (Kapa Biosystems, Woburn, MA, USA) in a final volume of 10  $\mu$ l (20 ng genomic DNA), using 3 mM  $MgCl_2$  and 0.2  $\mu$ M each primer. Products were purified by Exo-SAP digestion with exonuclease I (Epicentre, Madison, WI, USA) and shrimp alkaline phosphatase (GE Healthcare, Barcelona, Spain). Subsequently, single-base extension reactions were performed with the SNaPshot Multiplex kit (Applied Biosystems, Foster City, CA, USA). The genotyping call rate success of the newly studied SNPs was 99.12%. Sequences of primers and probes are available from the authors upon request.

### Selection of SNPs

Six ancestry informative markers (AIMs) were selected (see Table S1 in Additional file 1). Three of these are the most informative AIMs in differentiating Northern Europeans from Southern Europeans according to a large study [22]. Another two AIMs are the most informative for East-West place of origin according to the same study [22]. *rs12913832* is a SNP associated with large differences in frequency across Europe and unrelated to the previous [23]. In addition, we used genotype data from another 25 SNPs tagging 22 SLE loci reported in large European studies (see Table S1 in Additional file 1). These included nine SNPs in nine SLE loci we had already replicated and that made up the first phase of the current study [24]. We selected 14 additional SNPs for *de novo* genotyping in our samples. These SNPs were the top SLE-associated SNPs in large previous studies [2,3,7,27,28]. Not all of them have reached a genome-wide association level, but they are considered solid because they were found in large studies and with an odds ratio (OR) of the risk allele  $> 1.15$  in at least one study. These 14 SNPs together with two *IRF5* SNPs we had already studied [25] made up the 16 SNPs included in the second phase of our study.

### Statistical analysis

Analysis of results was based on R and Statistica 7.0 (StatSoft, Tulsa, OK, USA). Conformity with Hardy-Weinberg equilibrium was tested in control samples. Allele frequencies of the AIMs were compared between patients and controls from each collection with  $2 \times 2$  contingency tables. We created a global score by sample collection for the North-South axis of European population differentiation (N/S score) with the allele frequencies of the AIMs. First, we defined AIM allele frequencies as a function of the allele more common in Northern European populations. The most informative, nonredundant three AIMs were then selected. Finally, as a normalisation step we rescaled the frequencies from each of these three AIMs to 0 to 100%. This was done by considering 0% the frequency in the sample collection where it was less abundant and 100% the frequency where it was most abundant. The rescaled values of the three AIMs were averaged to obtain a combined normalised unique score for each collection.

Case-control allele frequencies were compared with fixed-effects and random-effects models stratifying by sample collection. For the fixed-effect model, the Mantel-Haenszel approach was used. For the random-effect model, an inverse variance meta-analysis approach was followed. Heterogeneity of effect sizes was evaluated with the inconsistency parameter  $I^2$  derived from the Cochran Q statistic. A high, moderate and low level of

**Table 1 Sample collections, proportion of women, frequencies of the three more informative AIMs and North/South score**

Collection <sup>a</sup>	DNA sample		Women (%)		AIM frequency (%)			N/S score
	Control	SLE	Control	SLE	rs12913832	rs382259	rs6730157	
NL	180	104	59.4	86.5	83.0	77.6	72.5	100.0
BE <sup>b</sup>	106	147	93.5	90.5				
DE	95	90	n.a.	90	77.9	68.1	56.9	80.5
CZ <sup>b</sup>	100	101	32	85.1				
SK	93	94	93.5	91.5	72.8	63.8	39.4	64.4
HU	95	95	48.4	90.4	63.4	59.6	39.0	54.9
ES, LCG	145	88	82.9	92	35.4	67.4	45.3	47.6
ES, OVD	200	147	69	91.8	30.5	70.4	45.4	47.2
ES, SCQ	95	109	48.4	96	35.1	68.2	41.3	46.0
ES, BCN	97	90	52.6	91.1	32.6	63.2	41.6	40.5
ES, MAD	281	92	68.8	91.3	33.1	61.9	42.3	40.0
PT	97	100	91.3	95.3	28.7	60.7	39.1	34.6
IT, MXP	106	129	42.5	86.8	47.6	49.1	19.8	26.2
IT, ROM	102	84	55.9	89.7	38.2	51.9	15.3	20.3
GR, AUMS	100	95	92	91.6	36.0	44.0	13.4	11.3
IT, NAP	109	79	100	90.1	38.6	42.2	10.0	9.6
GR, EH	100	98	67	86.7	36.7	38.0	13.5	6.8

Collections are in descending order of N/S score. <sup>a</sup>Collections from: University Medical Center Groningen, The Netherlands (NL); Université Catholique de Louvain, Belgium (BE); Hannover Medical School, Germany (DE); Institute of Rheumatology, The Czech Republic (CZ); Martin Faculty Hospital, Slovakia (SK); Albert Szent-Györgyi Medical and Pharmaceutical Centre, Hungary (HU); Complejo Hospitalario Universitario A Coruña, Spain (ES, LCG); Hospital Universitario Central de Asturias, Spain (ES, OVD); Hospital Clínico Universitario de Santiago, Spain (ES, SCQ); Hospital Val d'Hebron of Barcelona, Spain (ES, BCN); Hospital 12 de Octubre, Madrid, Spain (ES, MAD); Hospital Garcia de Orta, Almada, Portugal (PT); University of Milan and Fondazione IRCCS Ospedale Maggiore Policlinico from Milan, Italy (IT, MXP); Ospedale S. Camillo - Forlanini from Roma, Italy (IT, ROM); Athens University Medical School, Athens, Greece (GR, AUMS); Second University of Naples, Italy (IT, NAP); and Evangelismos Hospital, Athens, Greece (GR, EH). <sup>b</sup>These two collections were not analysed further because of differences in AIM frequencies between cases and controls.

inconsistency was attributed to levels of  $I^2$  over 75%, 50% and 25%, respectively, as described previously [29].

Distributions of ORs in Central European and Southern European populations were compared with the binomial distribution. Geometric mean ( $G_{\text{mean}}$ ) values of the ORs were obtained and compared. The sum of SLE risk alleles was obtained for the 20 SNPs showing independent association with SLE. The sum of genetic risk scores (GRSs) was also calculated for the same 20 SNPs. The total GRS for each patient with SLE was the sum of the products of the natural logarithm of the OR by the number of risk alleles at each locus that was carried by this patient. The ORs used to calculate GRS were the specific Mantel-Haenszel ORs for the corresponding European subgroup. The  $G_{\text{mean}}$  OR, mean of sum of risk alleles and mean cumulative GRS values were compared between groups with Student  $t$  tests. Correlation between the N/S score and sum of risk alleles or the mean of the natural logarithm of the OR was analysed with the weighted Pearson correlation coefficient. The threshold for significance was set at  $P \leq 0.05$ .

## Results

### Analysis of population differentiation

Our study included samples from 1,742 patients with SLE and 2,101 healthy controls recruited in 17 centres

(Table 1). Recruiters at each centre asked the patients and the controls for their ancestry, and only those reporting uniform known ancestry from the respective country were included. In addition, we checked with six top AIMs informative for European population differentiation whether there were differences between cases and controls from each recruitment centre. Five of the six AIMs provided completely independent information (pairwise  $r^2$  between them  $< 0.03$ ) - only rs6730157 and rs4988235 were redundant ( $r^2 = 0.85$ ). This analysis showed significant differences in the samples from two collections, the Czech Republic and Belgium. Consequently, these two collections of samples were discarded from all subsequent analyses.

Next, we divided the remaining 15 collections following the major axis of European population differentiation, the North-South axis. According to previous studies we expected two groups [20-22]: one with all samples from Central Europe, with samples from the Netherlands, Germany, Hungary and Slovakia (383 patients with SLE and 463 healthy controls); and a second with all samples from Portugal, Spain, Italy and Greece (1,111 patients and 1,432 controls). The AIM frequencies in samples from each collection were congruent with this division (Table 1). We obtained the

N/S score (for the place of each collection on the North-South axis), which was as expected from geographical distribution and previous studies [20-22] and was in agreement with the division we made (Table 1).

#### Nine SLE susceptibility loci with a Southern bias

We have already replicated association of top SNPs in nine SLE susceptibility loci in the samples included in this study (Table 2) [24]. When these SNPs were analysed separately in the two subgroups, Central European and Southern European, we found a bias for stronger association in the latter (Figure 1A). This bias was observed in eight of the nine SNPs. This distribution is significantly different from that expected by chance ( $P = 0.039$ ), and was observed with the Mantel-Haenszel OR (Figure 1A) and with the random-effect meta-analysis OR (see Figure S1 in Additional file 1). Mantel-Haenszel analysis was preferred because none of the 18 analyses showed high heterogeneity and only three showed a moderate level of

inconsistency. It should be noted that none of the SNPs taken individually was significantly different between the two groups.

#### Association analysis of additional SLE loci

We wanted to assess whether the bias found was a general phenomenon of SLE loci. We therefore selected 16 SNPs identifying other SLE genetic loci in Europeans [2,3,7,27,28]. All of the SNPs were genotyped successfully and all were in Hardy-Weinberg equilibrium when analysed by collection (the two *IRF5* SNPs have already been studied in a fraction of the samples [25]). The combined data showed significant differences between SLE cases and controls for 12 SNPs (Table 2). All of the significant differences were in the same direction as originally reported. Only four SNPs were similar in cases and controls, so they were excluded from further analysis. To avoid redundancy, we checked with conditional logistic regression whether the two associated SNPs in *IRF5*, *TNFAIP3*, *TNFSF4* or the two SNPs in the HLA (*rs2187668* and *rs3131379*) contributed independently to the association. One of the *TNFSF4* SNPs (*rs844644*) showed no association when conditioned in the other *TNFSF4* SNP ( $P = 0.072$ ), and therefore was no longer considered. On the contrary, the two *IRF5*, the two *TNFAIP3* and the two HLA SNPs remained associated and were included in the following analyses.

Stratification of the SLE patients and controls in Central Europeans and Southern Europeans showed that two of the 11 SNPs (*rs3131379* in *MSH5*,  $P = 0.003$ ; and *rs2187668* in *HLA-DQA1*,  $P = 0.046$ ) were significantly more associated in the Southern subgroup than in the Central European subgroup in the Mantel-Haenszel meta-analysis (Figure 1B); but none was significantly different in the random-effects meta-analysis (see Figure S2 in Additional file 1). The Mantel-Haenszel meta-analysis was favoured because none of the 22 analyses showed high inconsistency and only five showed a moderate level. In total, seven of the 11 associated SNPs were numerically more associated in the Southern European subgroup (Figure 1B). Only three SNPs were more associated in the Central Europeans (*rs573775* in *ATG5*, *rs729302* in *IRF5*, and *rs5754217* in *UBE2L3*) and one was equally associated in the two subgroups (*rs2205960* in *TNFSF4*), but none of these differences were significant.

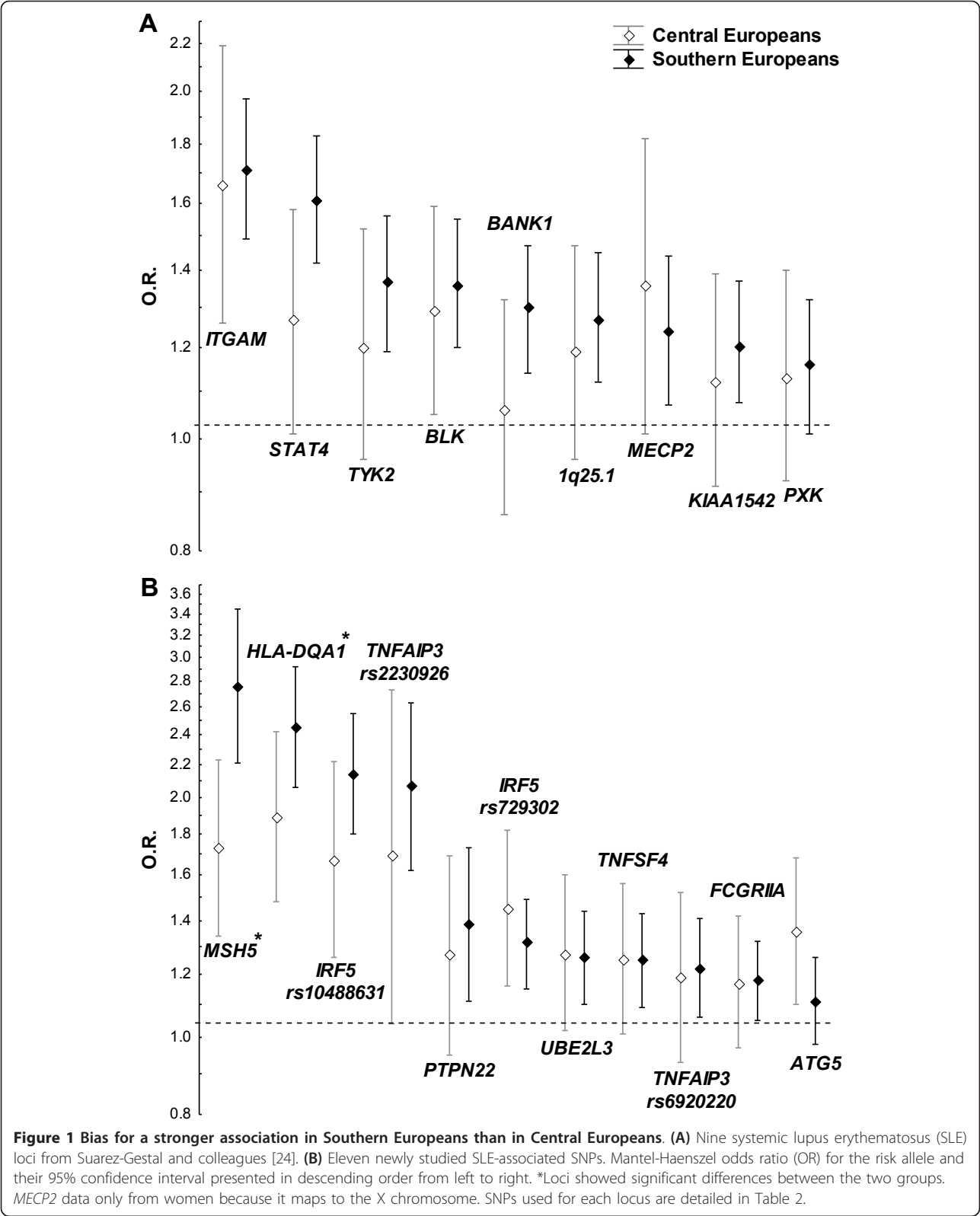
#### Southern bias for all of the SLE-associated SNPs together

When the 20 SNPs (nine from the first phase and 11 from the second) that have shown independent association in our samples were considered together, a bias towards a stronger association in the Southern subgroup was observed both as a significant deviation of the OR from a random binomial distribution ( $P = 0.019$ ) and as a significant difference between the OR means, which was larger

**Table 2 Association results of the 25 systemic lupus erythematosus SNPs**

SNP (locus)	Mantel-Haenszel analysis		Reference
	OR (95% CI)	P value	
<i>rs1143679</i> (ITGAM)	1.70 (1.50 to 1.93)	$5.16 \times 10^{-17}$	[24]
<i>rs7574865</i> (STAT4)	1.52 (1.36 to 1.70)	$6.11 \times 10^{-14}$	[24]
<i>rs13277113</i> (C8orf13-BLK)	1.34 (1.20 to 1.50)	$1.33 \times 10^{-7}$	[24]
<i>rs2304256</i> (TYK2)	1.32 (1.18 to 1.47)	$1.17 \times 10^{-6}$	[24]
<i>rs17435</i> (MECP2) <sup>a</sup>	1.27 (1.11 to 1.45)	$5.57 \times 10^{-4}$	[24]
<i>rs10798269</i> (1q25.1)	1.25 (1.11 to 1.39)	$6.12 \times 10^{-5}$	[24]
<i>rs17266594</i> (BANK1)	1.23 (1.10 to 1.37)	$1.92 \times 10^{-4}$	[24]
<i>rs4963128</i> (KIAA1542)	1.19 (1.08 to 1.32)	$1.31 \times 10^{-3}$	[24]
<i>rs6445975</i> (PXX)	1.15 (1.03 to 1.27)	0.02	[24]
<i>rs3131379</i> (MSH5)	2.25 (1.89 to 2.68)	$6.41 \times 10^{-20}$	-
<i>rs2187668</i> (HLA-DQA1)	2.17 (1.88 to 2.51)	$1.09 \times 10^{-25}$	-
<i>rs10488631</i> (IRF5)	2.00 (1.73 to 2.32)	$8.38 \times 10^{-21}$	[25]
<i>rs2230926</i> (TNFAIP3)	1.99 (1.60 to 2.47)	$2.54 \times 10^{-10}$	-
<i>rs729302</i> (IRF5)	1.35 (1.20 to 1.49)	$1.70 \times 10^{-7}$	[25]
<i>rs2476601</i> (PTPN22)	1.34 (1.13 to 1.60)	$9.15 \times 10^{-4}$	-
<i>rs5754217</i> (UBE2L3)	1.26 (1.13 to 1.42)	$7.30 \times 10^{-5}$	-
<i>rs2205960</i> (TNFSF4)	1.25 (1.11 to 1.40)	$1.32 \times 10^{-4}$	-
<i>rs6920220</i> (TNFAIP3)	1.21 (1.07 to 1.37)	$2.25 \times 10^{-3}$	-
<i>rs844644</i> (TNFSF4)	1.19 (1.08 to 1.32)	$7.22 \times 10^{-4}$	-
<i>rs1801274</i> (FCGR2A)	1.18 (1.07 to 1.30)	$1.27 \times 10^{-3}$	-
<i>rs573775</i> (ATG5)	1.17 (1.05 to 1.31)	$4.35 \times 10^{-3}$	-
<i>rs10156091</i> (ICA1)	1.09 (0.93 to 1.28)	0.3	-
<i>rs4240671</i> (XKR6)	1.09 (0.98 to 1.19)	0.1	-
<i>rs2667978</i> (LYN)	1.06 (0.94 to 1.19)	0.3	-
<i>rs6922466</i> (PERP)	1.02 (0.91 to 1.14)	0.8	-

The first nine SNPs were analysed in the first phase of the study and the remaining 16 in the second. Results of Mantel-Haenszel analysis are shown. All odds ratios (ORs) are expressed for the risk allele. Analyses of some of the SNPs in these samples have already been reported [24,25]. CI, confidence interval. <sup>a</sup>Analysis performed only in women because the locus is in the X chromosome.

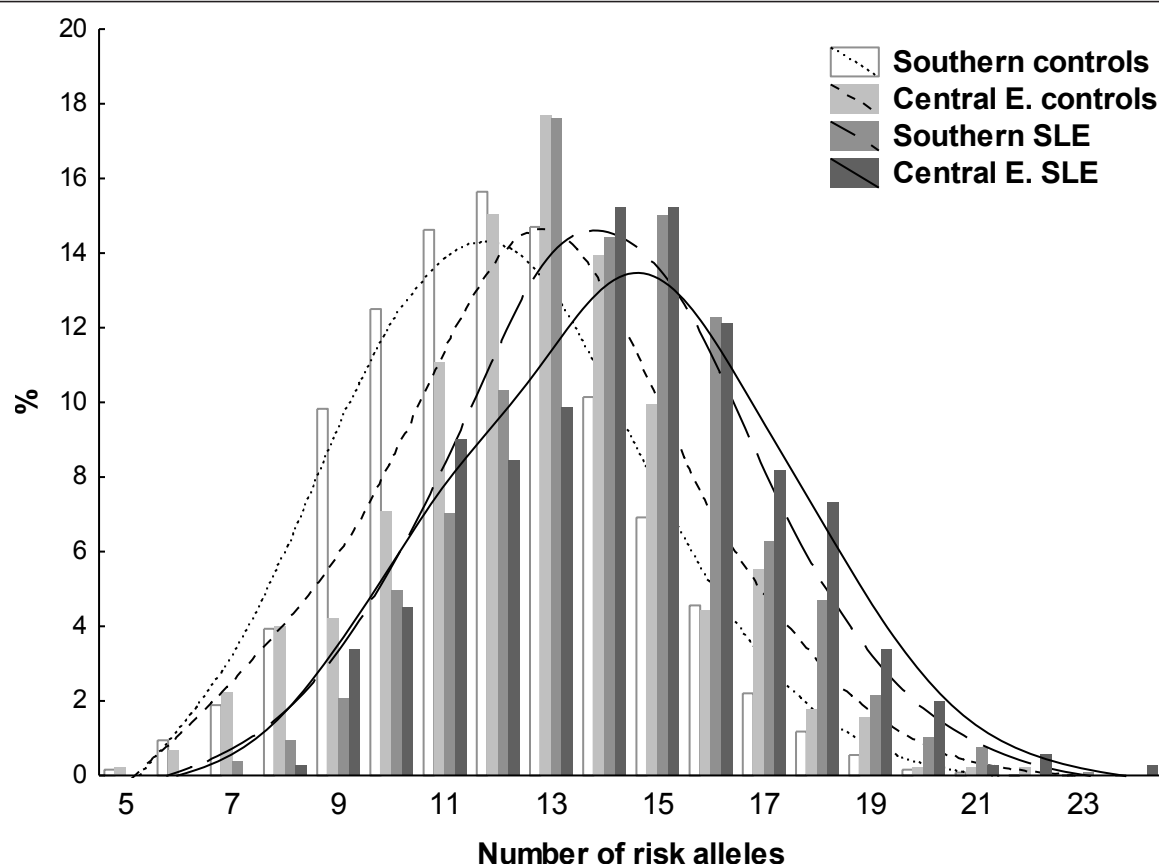


in the Southern samples ( $G_{\text{mean}} = 1.46 \pm 1.30$ ) than in the Central European samples ( $G_{\text{mean}} = 1.34 \pm 1.17$ ,  $P = 0.02$ ). Because it was possible that a fraction of the effect size attributed to a SNP is dependent on other SLE-associated SNPs, we also compared the mean of the OR for each SNP conditional on all the other SNPs. This analysis also showed a larger effect size in the Southern subjects ( $G_{\text{mean}} = 1.43 \pm 1.24$ ) than in the Central Europeans ( $G_{\text{mean}} = 1.28 \pm 1.18$ ,  $P = 0.02$ ).

We also analysed our data in a different way by counting the SLE risk alleles carried by each subject. Although it was possible to have from zero to 40 risk alleles, none of the subjects had less than five or more than 24 risk alleles. The distribution of frequencies stratified by disease status and by Central versus Southern subpopulations showed a gradient of values (Figure 2). The lowest number of risk alleles was observed in the healthy controls from the Southern European group (mean  $\pm$  standard deviation =  $12.0 \pm 2.5$ ). Immediately higher was the number of risk alleles corresponding to the Central European controls

( $12.8 \pm 2.7$ ,  $P = 1.1 \times 10^{-8}$  vs. the Southern European controls). This group was followed for the Southern European SLE patients ( $14.0 \pm 2.5$ ,  $P = 2.0 \times 10^{-17}$  vs. the Central European controls) and, finally, for the Central European patients ( $14.4 \pm 2.8$ ,  $P = 0.016$  vs. the Southern SLE patients).

The differences in number of risk alleles were not due to confounding by deviations from Hardy-Weinberg equilibrium in any of the four sample groups for any of the SNPs or by differences in call rate between the four groups (data not shown). We also checked that the differences and order hold when only women were analysed, with the lowest number in Southern controls ( $12.0 \pm 2.4$ ) followed by Central controls ( $13.0 \pm 2.6$ ), Southern SLE patients ( $14.1 \pm 2.5$ ) and Central European patients ( $14.4 \pm 2.8$ ) - showing significant differences between each of these groups except the last two ( $P = 5.3 \times 10^{-9}$ ,  $P = 2.1 \times 10^{-8}$  and  $P = 0.056$ , respectively). The same sequence was observed when the comparison was made in men (Southern controls



**Figure 2 Systemic lupus erythematosus risk alleles in Central and Southern European patients and controls.** Distribution of the number of systemic lupus erythematosus (SLE) risk alleles in Central European and Southern European SLE patients and controls. y axis, percentage of subjects in each of the four strata with the corresponding total number of SLE risk alleles. Distance-weighted least-squares fitting lines are presented. White bars and dotted line, Southern European controls; light grey bars and short-slashed line, Central European controls; medium grey bars and long-slashed line, Southern European SLE patients; dark grey bars and continuous line, Central European SLE patients.



$12.0 \pm 2.7$  and Central controls  $12.7 \pm 3.0$ ,  $P = 0.01$ ; Southern SLE patients  $13.5 \pm 2.5$ ,  $P = 0.036$ ; Central SLE patients  $14.7 \pm 2.8$ ,  $P = 0.027$ ). In this case, the differences between groups were all significant in spite of the small size of the SLE patient groups.

These results showed from a different perspective the same bias in effect sizes that has been described in the previous paragraphs, because the difference in number of risk alleles between SLE patients and controls of Southern origin (difference = 2.06, 95% confidence interval = 1.85 to 2.26) was significantly larger than for patients and controls of Central European origin (difference = 1.63, 95% confidence interval = 1.25 to 2.01;  $P = 0.027$ ). In addition, these results showed that the SLE risk alleles were less frequent in the Southern European subjects overall, but the difference was significantly more marked between controls from the Southern and Central subgroups (difference = 0.81, 95% confidence interval = 0.53 to 1.09) than between SLE patients from the same subgroups (difference = 0.38, 95% confidence interval = 0.07 to 0.69;  $P = 0.023$ ).

We also compared the cumulative GRS between Southern and Central European SLE patients. This parameter includes information from the sum of risk alleles and from the OR, and therefore is not independent of previous comparisons (Figure 3). The mean sum GRS was significantly larger in Southern European patients than in Central European patients ( $4.31 \pm 1.17$  vs.  $3.48 \pm 0.93$ ,  $P = 1.8 \times 10^{-32}$ ). The difference persisted after excluding the two HLA SLE-associated SNPs from the analysis ( $3.74 \pm 0.84$  vs.  $2.92 \pm 0.68$ ,  $P = 2.4 \times 10^{-58}$ ).

Finally, we wanted to analyse the mean number of SLE risk alleles for each sample collection as a function of its position along the North-South axis of population differentiation (Figure 4). This analysis showed that the number of SLE risk alleles in controls and in patients with SLE increased with the N/S score ( $R^2 = 0.67$ ,  $P = 0.002$  for controls; and  $R^2 = 0.45$ ,  $P = 0.002$  for patients). Similarly, the mean of the natural logarithm of OR correlated with the N/S score ( $R^2 = 0.40$ ,  $P = 0.012$ ). The gradual change in function of the score indicates that the previously performed analyses were not sensitive to the point used to separate Central European from Southern European populations.

#### Frequency clines of the SLE risk alleles

The previous results suggest the possibility of frequency gradients or clines of SLE risk alleles along the North-South axis of population differentiation. We therefore analysed this possibility for each of the SNPs. For this analysis we only used data from controls because they are more representative of the general population. None of the SLE-associated SNPs showed significant linkage disequilibrium with the AIMs included in this study (all

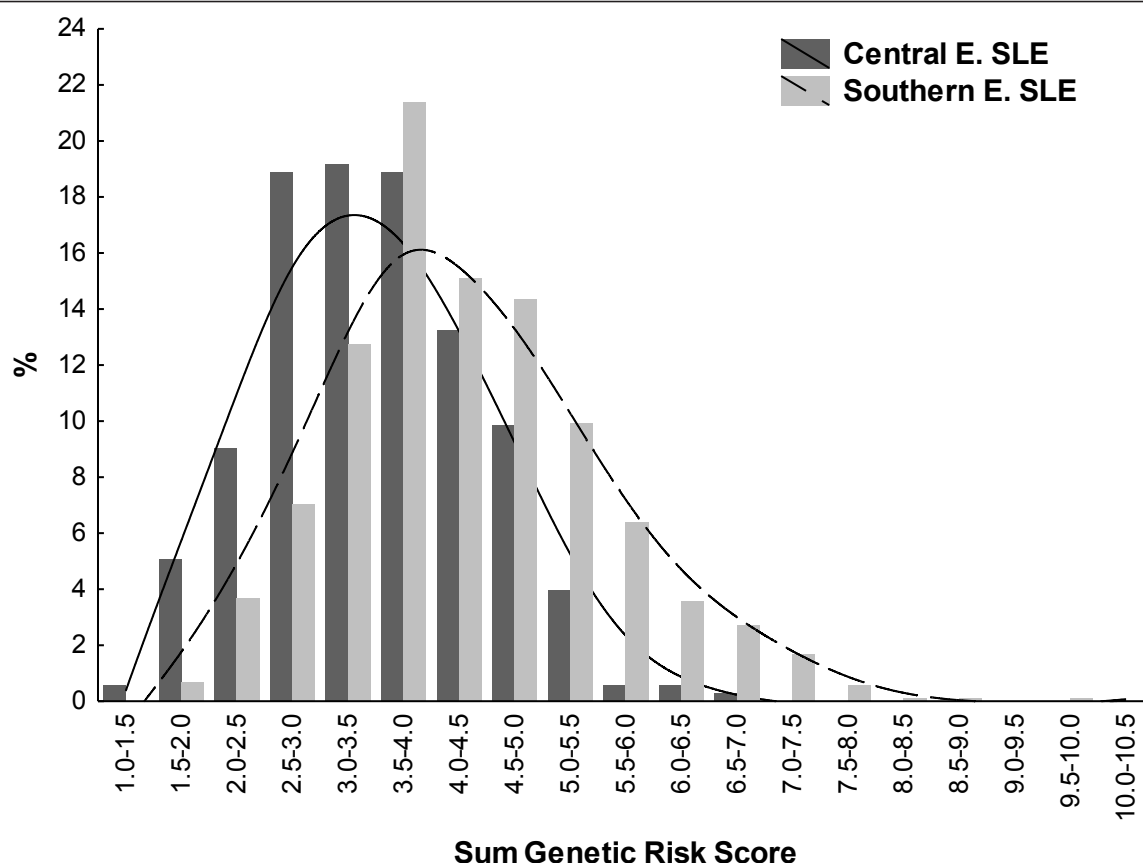
pairwise  $r^2$  between AIMs and SLE-associated SNPs < 0.05). However, 10 SNPs were significantly different between Central Europeans and Southern Europeans (Table 3). Five of these showed differences in excess of 5%, including the two SNPs in the *HLA* region, *PTPN22* (which is already known to show a cline in Europe [30]), *BLK* and *PXK*. Eight of the 10 SNPs were more common in controls from the Central group than from the Southern group in accordance with the direction of change observed with the sum of all risk alleles. The two exceptions were the *ITGAM* and *FCGR2A* SNPs. These two SNPs, however, have shown the same bias in effect sizes as the other eight. We also checked how the SNP frequencies in controls fitted a linear regression as a function of the N/S score. Results were similar to the obtained with the Central versus Southern group comparisons, except for SNPs with modest differences and for the SNP in *BLK* (Table 3).

#### Discussion

Two main aspects of our results should be highlighted: a significant bias to larger effect sizes of the SLE susceptibility loci in subjects from Southern Europe than those from Central Europe, and a lower frequency of the SLE risk alleles at these loci in subjects from Southern Europe.

The bias to stronger association among Southern Europeans was shown with four types of analysis. The first, comparing the number of SLE loci showing a trend to stronger associations beyond the expected at random, is independent of the loci characteristics. In contrast, comparison of the mean OR reflects the magnitude of the differences in effect size. An advantage of these two analyses is that they were done with meta-analysis approaches to account for sample collection factors. Their limitation is that they convey little information about the nature of the differences. The next analysis, comparison of the number of SLE risk alleles carried by each subject, is more informative but does not account for sample collection effects. The last analysis, comparison of sum GRS, combines information from the two previous analyses and is therefore not independent. Concordance of results from the different analyses is reassuring.

Further confidence was gained from the gradual change of the number of SLE risk alleles per sample collection as a function of the N/S score, implying insensitivity of the results to the specific partition of Europeans we have used. It is also important to note that the studies which identified SLE risk loci were carried out with subjects of a dominant Northern European ancestry (full references in Table S1 in Additional file 1), and therefore the stronger association we have found in Southern Europeans cannot be attributed to ascertainment bias. In other words, any bias due to the discovery of SLE loci in Northern-Central Europeans will favour a stronger effect size in that



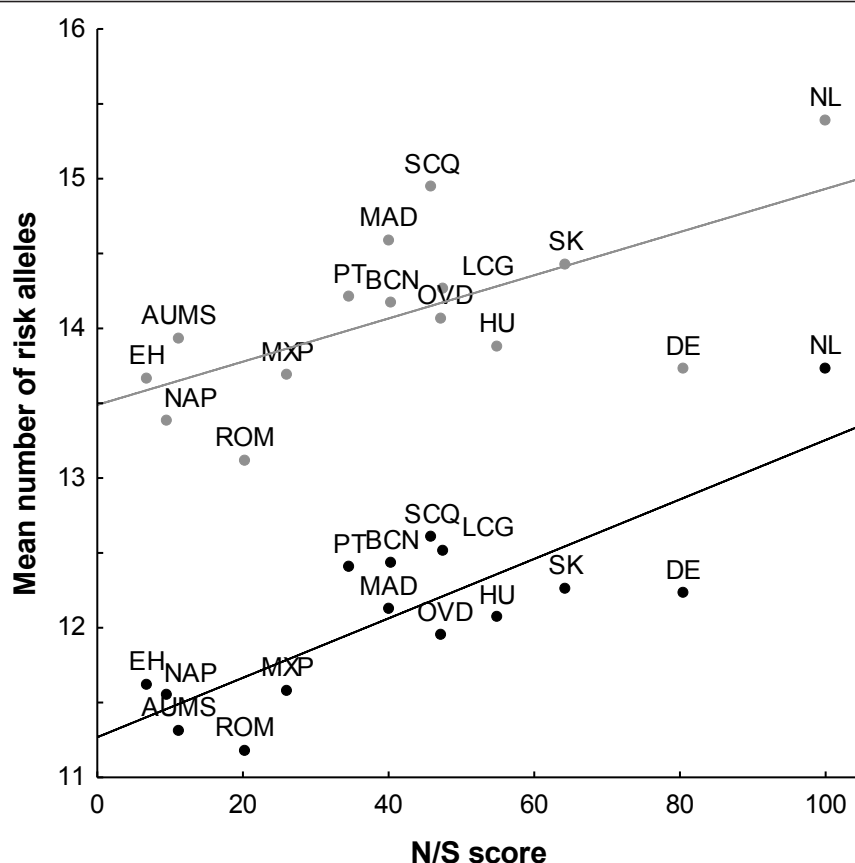
**Figure 3 Sum of genotypic risk score in Central and Southern European patients with systemic lupus erythematosus.** y axis, percentage of subjects in each of the two groups with sum genetic risk score (GRS) over 20 SNPs in the indicated intervals. Distance-weighted least-squares fitting lines are presented. Light grey bars and long-slashed line, Southern European systemic lupus erythematosus (SLE) patients; dark grey bars and continuous line, Central European SLE patients.

subpopulation, which is the opposite of our findings. This makes it very unlikely that our results are due to a tighter linkage disequilibrium between causal SNPs and the studied SNPs in the Southern subjects. All of these considerations support the validity of our findings. One should note that it was an average effect, however, because not all of the SLE risk alleles showed a trend to stronger effect sizes in the Southern samples and, in most of those that showed the trend, differences were small and not significant when assessed individually.

The second aspect of our results is the differential distribution of SLE risk alleles, with a lower frequency in Southern Europeans than in Central Europeans. The difference was observed both in SLE patients and in controls, but was more marked in the latter. It was maximal between the subjects from Greece and Italy, on one side, and those for the Netherlands on the other. All of the remaining groups were in between. Seven of the SNPs showed a frequency cline correlating with the N/S score. Only one of these clines has already been described for the risk allele of *PTPN22* [30].

The existence of allele frequency clines in the European population and their main axis of differentiation are well established [20-23]. What is surprising about our results is that the number of SLE risk alleles followed a gradient along this axis instead of varying randomly, with some alleles more common towards the North and others towards the South. This observation suggests the possible effect of selective forces acting along the history of the European populations, probably through resistance to infections [31,32]. These ideas demand new studies aiming to explore the relationship of autoimmunity with infection vulnerability.

A notable facet of the cline of SLE risk alleles in our study was that the difference between Southern Europeans and Central Europeans is less marked in SLE patients than in controls. We propose that this is due to the genetic structure of SLE that makes patients more similar at SLE loci across European subpopulations than the average member of the same subpopulations. This hypothesis is derived from the polygenic model of genetic inheritance [33], which includes the concept of liability threshold:



**Figure 4 Correlation between North-South population differentiation axis and number of systemic lupus erythematosus risk alleles.** Mean number of SLE risk alleles per sample collection is represented against the North/South score (N/S score) obtained in the same samples. Collection letter codes are as in Table 1.

disease appears when the contribution of multiple genetic factors to disease liability overcomes a threshold. Diseased subjects are therefore, on average, more similar for the genes involved in the disease than are the controls. More complex scenarios with involvement of differential environmental factors are also possible, however, and the risk allele of a particular locus could follow a different pattern because of specific factors besides the generic effect proposed here.

The validity of our results will be reinforced by replication in other sample collections. This will be particularly important in relation to the Central European subgroup - this was the smallest in our study, causing lower precision in the OR values and lower power to detect differential effects. Data from additional populations would also be interesting, in particular from the more extreme European subpopulations, because coverage of the full European spectrum would allow detecting additional clines. In addition, the AIMs we used were sufficient for group-level analyses but not for classification of individual subjects, which could have made our analyses more powerful. The

studied AIMs were able to show the European population substructure along the North-South differentiation, however, as in the studies where these AIMs were selected [22,23].

Our study also provides independent replication of SLE loci. Most of these loci were already strongly established and do not require comment (full references in Table S1 in Additional file 1). In contrast, four SNPs with previous solid association were not replicated in our analysis. Probably the most solid of them is *rs10156091* in *ICA1*. This SNP was first associated with SLE in a large genome-wide association scan (GWAS;  $P = 1.9 \times 10^{-7}$ , OR = 1.32) [3]. This result was confirmed in an even larger replication study, but with a lower effect size (OR = 1.16,  $P = 6.5 \times 10^{-4}$ ) [7]. The allelic frequency of this SNP in controls (10.0%) implies that our study had 97% power to detect the originally described effect with  $P < 0.05$ , but only 51% power for an OR like that observed in the replication study.

Also very solid is the record of *rs2667978* in *LYN*. This was discovered in the same GWAS ( $P = 5.1 \times 10^{-8}$ ,



**Table 3 Risk allele frequencies in Central and Southern European controls and relationship with the North/South score**

SNP (locus)	RAF (%)		$\chi^2$ P value	RAF vs. N/S score		
	Central European	Southern European		$r_{xy}$	Slope	P value
rs3131379 ( <i>MSH5</i> )	14.9	4.6	$4.9 \times 10^{-26}$	0.92	4.15	$2.0 \times 10^{-6}$
rs2187668 ( <i>HLA-DQA1</i> )	15.7	8.7	$1.9 \times 10^{-9}$	0.88	4.93	$4.0 \times 10^{-5}$
rs2476601 ( <i>PTPN22</i> )	11.8	6.1	$1.5 \times 10^{-8}$	0.74	4.21	$2.7 \times 10^{-3}$
rs13277113 ( <i>C8orf13-BLK</i> )	31.0	23.2	$2.2 \times 10^{-6}$	0.20	0.95	NS
rs6445975 ( <i>PXK</i> )	29.0	22.6	$7.5 \times 10^{-5}$	0.66	2.67	0.01
rs1143679 ( <i>ITGAM</i> )	11.9	16.7	$4.9 \times 10^{-4}$	-0.59	-4.79	0.03
rs2205960 ( <i>TNFSF4</i> )	25.2	20.7	$4.9 \times 10^{-3}$	0.15	1.08	NS
rs1801274 ( <i>FCGR2A</i> )	44.5	48.6	0.03	-0.40	-1.25	NS
rs4963128 ( <i>KIAA1542</i> )	69.1	65.3	0.03	0.58	2.83	0.03
rs10488631 ( <i>IRF5</i> )	11.6	9.3	0.046	0.46	3.64	0.10
rs17435 ( <i>MECP2</i> ) <sup>a</sup>	20.9	22.8	NS	-0.57	-2.84	0.03
rs573775 ( <i>ATG5</i> )	28.2	26.3	NS	0.51	4.10	0.06
rs7574865 ( <i>STAT4</i> )	23.3	23.8	NS	-0.51	-4.63	0.06
rs6920220 ( <i>TNFAIP3</i> )	19.3	17.8	NS	0.49	2.74	0.07
rs10798269 ( <i>1q25.1</i> )	66.2	69.0	NS	-0.17	-1.16	NS
rs5754217 ( <i>UBE2L3</i> )	22.8	20.8	NS	0.16	1.28	NS
rs2304256 ( <i>TYK2</i> )	73.1	71.0	NS	0.06	0.23	NS
rs2230926 ( <i>TNFAIP3</i> )	3.5	4.3	NS	-0.21	-3.42	NS
rs729302 ( <i>IRF5</i> )	67.8	69.0	NS	-0.00	-0.03	NS
rs17266594 ( <i>BANK1</i> )	71.2	70.9	NS	0.32	1.99	NS

Comparison of the risk allele frequencies (RAFTs) between Central and Southern European controls, and linear regression of RAFTs to the North/South score (N/S score) of population differentiation (from Table 1). NS, not significant. <sup>a</sup>Analysis performed only in women because the locus is in the X chromosome.

OR = 0.81) [3] and confirmed in a large replication study with European descent samples, but more weakly ( $P = 0.016$ ) [34]. However, this association was not confirmed in the already mentioned large replication study where a different SNP was used, *rs7829816* [7]. Our study was powered to detect an effect as that originally described with  $P < 0.01$ . The *rs4240671* SNP in *XKR6* is also backed by strong evidence. It was identified in the same SLE GWAS as *ICA1* and *LYN* with very strong evidence including five SNPs with  $P < 5.0 \times 10^{-8}$  [3]. The association was not uniformly observed across sample collections in this study, however, and no replication in any large study has been reported. The only independent replication was obtained for one of the SNPs in a 245-family study in Canadians (*rs6985109*;  $P = 0.008$ ) [35]. Our study had enough power to detect the originally reported association (OR = 0.75) with  $P < 10^{-7}$  given its elevated allele frequency (49%). Finally, the weakest previous support was for *rs6922466* in *PERP*. This SNP was associated with SLE ( $P = 1.0 \times 10^{-4}$ ) only in a large study [28].

## Conclusion

Our study has uncovered a bias for stronger effect sizes of SLE risk alleles in Southern Europeans than in Central Europeans. This bias was accompanied by a lower frequency of the risk alleles in the Southern European

group. Difference in frequencies was more marked in controls than in patients with SLE. These results should be taken into account for genetic studies of SLE and for understanding the genetic structure of the disease and the possible presence of autoimmune disease risk allele clines and their causes. In addition, these results call for exploration of the assumptions and implications of the liability threshold concept - in particular, whether a constant threshold is consistent with GWAS data - as well as for exploration of environmental or other factors that could explain the effect size bias. Our findings therefore contribute to define the genetic epidemiology of SLE and suggest new lines of research for understanding the deep genetic structure of SLE.

## Additional material

**Additional file 1: Table S1 presenting a detailed description of all the SNPs included in the study, and Figures S1 and S2 with random effect meta-analysis results corresponding to the same data as Figure 1 in the manuscript.**

## Abbreviations

AIM: ancestry informative marker; CI: confidence interval;  $G_{mean}$ : geometric mean; GRS: genetic risk score; GWAS: genome-wide association scan; N/S score: North/South score; OR: odds ratio; PCR: polymerase chain reaction; SLE: systemic lupus erythematosus; SNP: single nucleotide polymorphism.

## Acknowledgements

EA-P is the recipient of a predoctoral bursary from the Instituto de Salud Carlos III (Spain). This work was supported by Fondo de Investigación Sanitaria of the Instituto de Salud Carlos III (Spain) grant 08/0744 and by RETICS Program RD08/0075 that are partially financed by the European Regional Development Fund of the European Union, and by grants from the Xunta de Galicia and BMBF KN Rheuma (C2.12 to TW). Other contributors to the European Consortium of SLE DNA Collections: Myriam Liz, Laboratorio de Investigación 10 and Rheumatology Unit, Instituto de Investigación Sanitaria - Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain; Reinhold E Schmidt, Division of Clinical Immunology, Department of Internal Medicine of the Hannover Medical School, Hannover, Germany; Iris Kappou-Rigatou, Department of Histocompatibility and Immunology, Evangelismos Hospital, Athens, Greece; Raffaella Scorza, Clinical Immunology, University of Milan and Fondazione IRCCS Ospedale Maggiore Policlinico, Mangiagalli e Regina Elena, Milan, Italy; Emöke Endereffy, Department of Paediatrics, Albert Szent-Györgyi Clinical Centre, University of Szeged, Szeged, Hungary; Eva Balada, Internal Medicine, Research Laboratory in Autoimmune Diseases Hospital Vall d'Hebron, Barcelona, Spain; Cees G Kallenberg, Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, Groningen, the Netherlands; Filipe Vinagre, Rheumatology Department, Hospital Garcia de Orta, Almada (Portugal) and Rheumatology Research Unit, Instituto Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Portugal; Ctibor Dostal, Molecular Biology and Immunogenetics Department, Institute of Rheumatology, Prague, Czech Republic; Rudolf Pullmann Jr, Gerontology Research Center, National Institute on Aging, Nathan Shock Drive, Baltimore, Maryland, USA; Maria Mavromati, Pathophysiology Department, Athens University Medical School, Athens, Greece; Sandra D'Alfonso, Department of Medical Sciences and IRCAD, Eastern Piedmont University, Novara, Italy; Carmen Gutierrez, Department of Functional Biology, Hospital Universitario Central de Asturias, Universidad de Oviedo, Oviedo, Spain; and Ignacio Rego, Laboratorio de Investigación Osteoarticular y del Envejecimiento, Servicio de Reumatología, CH Universitario A Coruña, A Coruña, Spain.

## Author details

<sup>1</sup>Laboratorio de Investigación 10 and Rheumatology Unit, Instituto de Investigación Sanitaria - Hospital Clínico Universitario de Santiago, Choupana s/n, Santiago de Compostela 15706, Spain. <sup>2</sup>U.O. Complessa di Reumatologia, Azienda Ospedaliera San Camillo - Forlanini, Piazza Carlo Forlanini 1, Rome 00151, Italy. <sup>3</sup>Institute of Clinical Biochemistry, Martin Faculty Hospital and Jessenius Medical Faculty, Kollárova 2, Martin 036 59, Slovakia. <sup>4</sup>Department of Histocompatibility and Immunology, Evangelismos Hospital, Ipsilantou Str. 45-47, Athens 10675, Greece. <sup>5</sup>Department of Rheumatology, Hospital of Hungarian Railways, Lukács György utca 4, Szolnok 5000, Hungary. <sup>6</sup>Pathophysiology Department, Athens University Medical School, Korai 5, Athens 115 27, Greece. <sup>7</sup>Department of Internal Medicine and Rheumatology, Martini Hospital, Van Swietenplein 1, Groningen 9728, the Netherlands. <sup>8</sup>Department of Functional Biology, University of Oviedo, Calle Catedrático Valentín Andrés Álvarez s/n, Oviedo 33006, Spain. <sup>9</sup>Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico and University of Milan, Via Francesco Sforza 35, Milan 20122, Italy. <sup>10</sup>Rheumatology Unit, Second University of Naples, Via Sergio Pansini 5, Naples 80131, Italy. <sup>11</sup>Rheumatology Department, Hospital 12 de Octubre, Av. Andalucía s/n, Madrid 28041, Spain. <sup>12</sup>Internal Medicine and Research Laboratory in Autoimmune Diseases, Hospital Vall d'Hebron, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. <sup>13</sup>Department of Internal Medicine and Division of Clinical Immunology, Hannover Medical School, Carl-Neuberg-Straße 1, Hannover 30625, Germany. <sup>14</sup>Institute of Biotechnology, Academy of Sciences of the Czech Republic, Národní 1009/3, Prague 110 00, Czech Republic. <sup>15</sup>Rheumatology Department, Hospital Garcia de Orta and Rheumatology Research Unit, Instituto de Medicina Molecular, Av. Torrado da Silva s/n, Lisboa 2801-951, Portugal. <sup>16</sup>Department of Medical Sciences and IRCAD, Eastern Piedmont University, Via dei Torielli 12, Novara 28100, Italy. <sup>17</sup>Laboratorio de Investigación Osteoarticular y del Envejecimiento, Servicio de Reumatología, INIBIC-CH Universitario, Jubas de Arriba 84A, A Coruña 15006, Spain. <sup>18</sup>Cliniques Universitaires Saint-Luc, Université catholique de Louvain, Avenue Emmanuel Mounier 81, Brussels 1200, Belgium. <sup>19</sup>Department of Medicine, University of Santiago de Compostela, Calle Choupana s/n, Santiago de Compostela 15706, Spain.

## Authors' contributions

EA-P and MS-G participated in the design of the study, genotyped the samples, and participated in the interpretation of the results and writing the manuscript. MC participated in the statistical analysis and in the interpretation of results. GDS, RP, CP, AK, FNS, MB, AS, MM, SM, PC, JO-R, TW, SR, MJS, NB, FJB, BRL and JIG-R participated in the acquisition of clinical data and collection of samples and in the analysis and interpretation of results. AG participated in the design of the study and in coordination of the acquisition of clinical data and collection of samples, and supervised laboratory work and interpretation and writing of the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 16 February 2012 Revised: 27 March 2012

Accepted: 27 April 2012 Published: 27 April 2012

## References

- Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, Hirankarn N, Ying D, Pan HF, Mok CC, Chan TM, Wong RW, Lee KW, Mok MY, Wong SN, Leung AM, Li XP, Avihingsanon Y, Wong CM, Lee TL, Ho MH, Lee PP, Chang YK, Li PH, Li RJ, Zhang L, Wong WH, Ng IO, Lau CS, Sham PC, et al: **Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus.** *PLoS Genet* 2010, **6**:e1000841.
- Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapaa-Dahlqvist S, Petri M, Manzi S, Seldin MF, Ronnblom L, Syvanen AC, Criswell LA, Gregersen PK, Behrens TW: **Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX.** *N Engl J Med* 2008, **358**:900-909.
- Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, Ojwang JO, et al: **Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci.** *Nat Genet* 2008, **40**:204-210.
- Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jonsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrop H, Gonzalez-Escribano MF, Martin J, Abderrahim H, Alarcon-Riquelme ME: **Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40**:211-216.
- Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, Xu JH, Cai ZM, Huang W, Zhao GP, Xie HF, Fang H, Lu QJ, Li XP, Pan YF, Deng DQ, Zeng FQ, Ye ZZ, Zhang XY, Wang QW, Hao F, Ma L, Zuo XB, Zhou FS, Du WH, Cheng YL, Yang JQ, Shen SK, Li J, Sheng YJ, et al: **Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus.** *Nat Genet* 2009, **41**:1234-1237.
- Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, Burtt NP, Guiducci C, Parkin M, Gates C, Plenge RM, Behrens TW, Wither JE, Rioux JD, Fortin PR, Graham DC, Wong AK, Vyse TJ, Daly MJ, Altshuler D, Moser KL, Gaffney PM: **Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40**:1059-1061.
- Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jonsen A, Bengtsson AA, Rantapaa-Dahlqvist S, Baechler EC, Brown EE, Alarcon GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vila LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, et al: **A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus.** *Nat Genet* 2009, **41**:1228-1233.
- Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, Pan HF, Mok CC, Chan TM, Wong RW, Mok MY, Lee KW, Wong SN, Leung AM, Li XP, Avihingsanon Y, Rianthavorn P, Deekajorndej T, Suphapeetiporn K, Shotelersuk V, Baum L,

- Kwan P, Lee TL, Ho MH, Lee PP, Wong WH, Zeng S, Zhang J, Wong CM, Ng IO, Garcia-Barcelo MM, et al: **ELF1 is associated with systemic lupus erythematosus in Asian populations.** *Hum Mol Genet* 2011, **20**:601-607.
9. Zhao J, Wu H, Khosravi M, Cui H, Qian X, Kelly JA, Kaufman KM, Langefeld CD, Williams AH, Comeau ME, Ziegler JT, Marion MC, Adler A, Glenn SB, Alarcon-Riquelme ME, Pons-Estel BA, Harley JB, Bae SC, Bang SY, Cho SK, Jacob CO, Vyse TJ, Niewold TB, Gaffney PM, Moser KL, Kimberly RP, Edberg JC, Brown EE, Alarcon GS, Petri MA, et al: **Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility.** *PLoS Genet* 2011, **7**:e1002079.
10. Sanchez E, Comeau ME, Freedman BI, Kelly JA, Kaufman KM, Langefeld CD, Brown EE, Alarcon GS, Kimberly RP, Edberg JC, Ramsey-Goldman R, Petri M, Reveille JD, Vila LM, Merrill JT, Tsao BP, Kamen DL, Gilkeson GS, James JA, Vyse TJ, Gaffney PM, Jacob CO, Niewold TB, Richardson BC, Harley JB, Alarcon-Riquelme ME, Sawalha AH: **Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study.** *Arthritis Rheum* 2011, **63**:3493-3501.
11. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, Adler A, Williams AH, Gallant CJ, Anaya JM, Bae SC, Boackle SA, Brown EE, Chang DM, Criswell LA, Edberg JC, Freedman BI, Gregersen PK, Gilkeson GS, Jacob CO, James JA, Kamen DL, Kimberly RP, Martin J, Merrill JT, Niewold TB, Park SY, Petri MA, Pons-Estel BA, Ramsey-Goldman R, Reveille JD, et al: **Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study.** *Am J Hum Genet* 2011, **88**:83-91.
12. Han S, Kim-Howard X, Deshmukh H, Kamatani Y, Viswanathan P, Guthridge JM, Thomas K, Kaufman KM, Ojwang J, Rojas-Villarraga A, Baca V, Orozco L, Rhodes B, Choi CB, Gregersen PK, Merrill JT, James JA, Gaffney PM, Moser KL, Jacob CO, Kimberly RP, Harley JB, Bae SC, Anaya JM, Alarcon-Riquelme ME, Matsuda K, Vyse TJ, Nath SK: **Evaluation of imputation-based association in and around the integrin- $\alpha$ -M (ITGAM) gene and replication of robust association between a non-synonymous functional variant within ITGAM and systemic lupus erythematosus (SLE).** *Hum Mol Genet* 2009, **18**:1171-1180.
13. Fu J, Festen EA, Wijmenga C: **Multi-ethnic studies in complex traits.** *Hum Mol Genet* 2011, **20**:R206-R213.
14. Sanchez E, Webb RD, Rasmussen A, Kelly JA, Riba L, Kaufman KM, Garcia-de la Torre I, Moctezuma JF, Maradiaga-Cecena MA, Cardiel-Rios MH, Acevedo E, Cucho-Venegas M, Garcia MA, Gamron S, Pons-Estel BA, Vasconcelos C, Martin J, Tusie-Luna T, Harley JB, Richardson B, Sawalha AH, Alarcon-Riquelme ME: **Genetically determined Amerindian ancestry correlates with increased frequency of risk alleles for systemic lupus erythematosus.** *Arthritis Rheum* 2010, **62**:3722-3729.
15. Chung SA, Tian C, Taylor KE, Lee AT, Ortmann WA, Hom G, Graham RR, Nititham J, Kelly JA, Morrisey J, Wu H, Yin H, Alarcon-Riquelme ME, Tsao BP, Harley JB, Gaffney PM, Moser KL, Manzi S, Petri M, Gregersen PK, Langefeld CD, Behrens TW, Seldin MF, Criswell LA: **European population substructure is associated with mucocutaneous manifestations and autoantibody production in systemic lupus erythematosus.** *Arthritis Rheum* 2009, **60**:2448-2456.
16. Richman IB, Chung SA, Taylor KE, Kosoy R, Tian C, Ortmann WA, Nititham J, Lee AT, Rutman S, Petri M, Manzi S, Behrens TW, Gregersen PK, Seldin MF, Criswell LA: **European population substructure correlates with systemic lupus erythematosus endophenotypes in North Americans of European descent.** *Genes Immun* 2010, **11**:515-521.
17. Alonso-Perez E, Suarez-Gestal M, Calaza M, Witte T, Papasteriades C, Marchini M, Migliaresi S, Kovacs A, Ordi-Ros J, Bijl M, Santos MJ, Ruzickova S, Pullmann R, Carreira P, Skopouli FN, D'Alfonso S, Sebastiani GD, Suarez A, Blanco FJ, Gomez-Reino JJ, Gonzalez A: **Association of systemic lupus erythematosus clinical features with European population genetic substructure.** *PLoS One* 2011, **6**:e29033.
18. Ferreiros-Vidal I, Gomez-Reino JJ, Barros F, Carracedo A, Carreira P, Gonzalez-Escribano F, Liz M, Martin J, Ordi J, Vicario JL, Gonzalez A: **Association of PDCD1 with susceptibility to systemic lupus erythematosus: evidence of population-specific effects.** *Arthritis Rheum* 2004, **50**:2590-2597.
19. Ferreiros-Vidal I, D'Alfonso S, Papasteriades C, Skopouli FN, Marchini M, Scorza R, Migliaresi S, Sebastiani GD, Endreffy E, Mavromati M, Kappou-Rigatou I, Ruzickova S, Dostal C, Schmidt RE, Witte T, Gomez-Reino JJ, Gonzalez A: **Bias in association studies of systemic lupus erythematosus susceptibility due to geographical variation in the frequency of a programmed cell death 1 polymorphism across Europe.** *Genes Immun* 2007, **8**:138-146.
20. Seldin MF, Shigeta R, Villoslada P, Selmi C, Tuomilehto J, Silva G, Belmont JW, Klareskog L, Gregersen PK: **European population substructure: clustering of northern and southern populations.** *PLoS Genet* 2006, **2**:e143.
21. Tian C, Kosoy R, Nassir R, Lee A, Villoslada P, Klareskog L, Hammarstrom L, Garchon HJ, Pulver AE, Ransom M, Gregersen PK, Seldin MF: **European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups.** *Mol Med* 2009, **15**:371-383.
22. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, Selmi C, Klareskog L, Pulver AE, Qi L, Gregersen PK, Seldin MF: **Analysis and application of European genetic substructure using 300 K SNP information.** *PLoS Genet* 2008, **4**:e4.
23. Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, Hankinson SE, Hu FB, Duffy DL, Zhao ZZ, Martin NG, Montgomery GW, Hayward NK, Thomas G, Hoover RN, Chanock S, Hunter DJ: **A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation.** *PLoS Genet* 2008, **4**:e1000074.
24. Suarez-Gestal M, Calaza M, Endreffy E, Pullmann R, Ordi-Ros J, Domenico Sebastiani G, Ruzickova S, Jose Santos M, Papasteriades C, Marchini M, Skopouli FN, Suarez A, Blanco FJ, D'Alfonso S, Bijl M, Carreira P, Witte T, Migliaresi S, Gomez-Reino JJ, Gonzalez A: **Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study.** *Arthritis Res Ther* 2009, **11**:R69.
25. Ferreira-Neira I, Calaza M, Alonso-Perez E, Marchini M, Scorza R, Sebastiani GD, Blanco FJ, Rego I, Pullmann R Jr, Pullmann R, Kallenberg CG, Bijl M, Skopouli FN, Mavromati M, Migliaresi S, Barizzone N, Ruzickova S, Dostal C, Schmidt RE, Witte T, Papasteriades C, Kappou-Rigatou I, Endreffy E, Kovacs A, Ordi-Ros J, Balada E, Carreira P, Gomez-Reino JJ, Gonzalez A: **Opposed independent effects and epistasis in the complex association of IRF5 to SLE.** *Genes Immun* 2007, **8**:429-438.
26. Hochberg MC: **Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus.** *Arthritis Rheum* 1997, **40**:1725.
27. Cunningham Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, Moser KL, Rioux JD, Altschuler D, Behrens TW, Vyse TJ: **Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus.** *Nat Genet* 2008, **40**:83-89.
28. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, Shifrin N, Petri MA, Kamboh MI, Manzi S, Seldin MF, Gregersen PK, Behrens TW, Ma A, Kwok PY, Criswell LA: **Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus.** *Nat Genet* 2008, **40**:1062-1064.
29. Higgins JP, Thompson SG, Deeks JJ, Altman DG: **Measuring inconsistency in meta-analyses.** *BMJ* 2003, **327**:557-560.
30. Gregersen PK, Lee HS, Batliwalla F, Begovich AB: **PTPN22: setting thresholds for autoimmunity.** *Semin Immunol* 2006, **18**:214-223.
31. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R: **Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution.** *PLoS Genet* 2011, **7**:e1002355.
32. Barreiro LB, Quintana-Murci L: **From evolutionary genetics to human immunology: how selection shapes host defence genes.** *Nat Rev Genet* 2010, **11**:17-30.
33. Cummings M: *Human Heredity: Principles and Issues* Belmont: Brooks/Cole; 2010.
34. Lu R, Vidal GS, Kelly JA, Delgado-Vega AM, Howard XK, Macwana SR, Dominguez N, Klein W, Burrell C, Harley IT, Kaufman KM, Bruner GR, Moser KL, Gaffney PM, Gilkeson GS, Wakeland EK, Li QZ, Langefeld CD, Marion MC, Divers J, Alarcon GS, Brown EE, Kimberly RP, Edberg JC, Ramsey-Goldman R, Reveille JD, McGwin G Jr, Vila LM, Petri MA, Bae SC, et al: **Genetic associations of LYN with systemic lupus erythematosus.** *Genes Immun* 2009, **10**:397-403.
35. Budarf ML, Goyette P, Boucher G, Lian J, Graham RR, Claudio JO, Hudson T, Gladman D, Clarke AE, Pope JE, Peschken C, Smith CD, Hanly J, Rich E, Boire G, Barr SG, Zummer M, Fortin PR, Wither J, Rioux JD: **A targeted association study in systemic lupus erythematosus identifies multiple susceptibility alleles.** *Genes Immun* 2011, **12**:51-58.

doi:10.1186/ar3818

**Cite this article as:** Alonso-Perez *et al.*: Bias in effect size of systemic lupus erythematosus susceptibility loci across Europe: a case-control study. *Arthritis Research & Therapy* 2012 **14**:R94.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



**Table S1:** SNPs included in the study with indication of their alleles, chromosome, locus, location in the locus, reference from which they were obtained and their use: phase 1 includes the 9 SLE associated SNPs from Suarez-Gestal et al. [24]; phase 2 the SNPs newly analyzed in this report; and AIM the 6 European Ancestry Informative Markers

SNP	Alleles <sup>a</sup>	Chr	Locus	Location	Reference	Use
<i>rs1143679</i>	G/A	16	<i>ITGAM</i>	Exon 3	1	phase 1
<i>rs7574865</i>	G/T	2	<i>STAT4</i>	Intron 3	2 - 7	phase 1
<i>rs13277113</i>	G/A	8	<i>C8orf13-BLK</i>	Intergenic	3	phase 1
<i>rs2304256</i>	C/A	19	<i>TYK2</i>	Exon 8	8, 9	phase 1
<i>rs17435</i>	A/T	X	<i>MECP2</i>	Intron 2	10	phase 1
<i>rs10798269</i>	G/A	1	<i>Iq25.1</i>	Intergenic	2	phase 1
<i>rs17266594</i>	T/C	4	<i>BANK1</i>	Intron 1	11	phase 1
<i>rs4963128</i>	G/A	11	<i>KIAA1542</i>	Intron 4	2	phase 1
<i>rs6445975</i>	T/G	3	<i>PXK</i>	Intron 5	2	phase 1
<i>rs509749</i>	A/G	1	<i>LY9</i>	Exon 8	12	phase 1
<i>rs3131379</i>	G/A	6	<i>MSH5</i>	Ittron 1	2	phase 2
<i>rs2187668</i>	G/A	6	<i>HLA-DQA1</i>	Intergenic	3	phase 2
<i>rs2230926</i>	T/G	6	<i>TNFAIP3</i>	Exon 3	13	phase 2
<i>rs6920220</i>	G/A	6	<i>TNFAIP3</i>	Intergenic	14	phase 2
<i>rs6922466</i>	A/G	6	<i>PERP</i>	Intergenic	13	phase 2
<i>rs2476601</i>	G/A	1	<i>PTPN22</i>	Exon 8	2, 14	phase 2
<i>rs844644</i>	C/A	1	<i>TNFSF4</i>	Intergenic	15	phase 2
<i>rs2205960</i>	G/T	1	<i>TNFSF4</i>	Intergenic	14, 15	phase 2
<i>rs1801274</i>	G/A	1	<i>FCGR1IA</i>	Exon 4	2, 14	phase 2
<i>rs10156091</i>	C/T	7	<i>ICA1</i>	Intron 5	2, 14	phase 2
<i>rs4240671</i>	A/G	8	<i>XKR6</i>	Intron 1	2	phase 2
<i>rs2667978</i>	A/G	8	<i>LYN</i>	Intron 8	2	phase 2
<i>rs5754217</i>	G/T	22	<i>UBE2L3</i>	Intron 1	2, 14	phase 2
<i>rs573775</i>	G/A	6	<i>ATG5</i>	Intron 7	2	phase 2
<i>rs6730157</i>	A/G	2	<i>RAB3GAP1</i>	Intron 22	16	AIM
<i>rs382259</i>	T/C	6	<i>Intergenic</i>	Intergenic	16	AIM
<i>rs12203592</i>	C/T	6	<i>IRF4</i>	Intron 4	16	AIM
<i>rs354690</i>	T/C	2	<i>Intergenic</i>	Intergenic	16	AIM
<i>rs4988235</i> <sup>b</sup>	C/T	2	<i>MCM6</i>	Intron 13	17	AIM
<i>rs12913832</i>	A/G	15	<i>HERC2</i>	Intron 87	18	AIM

<sup>a</sup> Major allele/minor allele

<sup>b</sup> This SNP was genotyped in 81.5 % of the samples and it was largely redundant with *rs6730157* ( $R^2 = 0.84$ )

#### References for Table S1:

1. Nath SK, Han S, Kim-Howard X, Kelly JA, Viswanathan P, Gilkeson GS, Chen W, Zhu C, McEver RP, Kimberly RP, Alarcón-Riquelme ME, Vyse TJ, Li QZ, Wakeland EK, Merrill JT, James JA, Kaufman KM, Guthridge JM, Harley JB: A nonsynonymous functional variante in integrin-alpha(M) (encoded by

ITGAM) is associated with systemic lupus erythematosus. *Nat Genet* 2008, 40:152-154.

2. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD, Nath SK, Guthridge JM, Cobb BL, Mirel DB, Marion MC, Williams AH, Divers J, Wang W, Frank SG, Namjou B, Gabriel SB, Lee AT, Gregersen PK, Behrens TW, Taylor KE, Fernando M, Zidovetzki R, Gaffney PM, Edberg JC, Rioux JD, et al.: *Genomewide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci*. *Nat Genet* 2008, 40:204-210.
3. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, Lee AT, Chung SA, Ferreira RC, Pant PV, Ballinger DG, Kosoy R, Demirci FY, Kamboh MI, Kao AH, Tian C, Gunnarsson I, Bengtsson AA, Rantapää-Dahlqvist S, Petri M, Manzi S, Seldin MF, Rönblom L, Syvänen AC, Criswell LA, Gregersen PK, Behrens TW: *Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX*. *N Engl J Med* 2008, 358:900-909.
4. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, de Bakker PI, Le JM, Lee HS, Batliwalla F, Li W, Masters SL, Booty MG, Carulli JP, Padyukov L, Alfredsson L, Klareskog L, Chen WV, Amos CI, Criswell LA, Seldin MF, Kastner DL, Gregersen PK: *STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus*. *N Engl J Med* 2007, 357:977-986.
5. Kobayashi S, Ikari K, Kaneko H, Kochi Y, Yamamoto K, Shimane K, Nakamura Y, Toyama Y, Mochizuki T, Tsukahara S, Kawaguchi Y, Terai C, Hara M, Tomatsu T, Yamanaka H, Horiuchi T, Tao K, Yasutomo K, Hamada D, Yasui N, Inoue H, Itakura M, Okamoto H, Kamatani N, Momohara S: *Association of STAT4 with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population*. *Arthritis Rheum* 2008, 58:1940-1946.
6. Palomino-Morales RJ, Rojas-Villarraga A, González CI, Ramírez G, Anaya JM, Martín J: *STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians*. *Genes Immun* 2008, 9:379-382.
7. Taylor KE, Remmers EF, Lee AT, Ortmann WA, Plenge RM, Tian C, Chung SA, Nititham J, Hom G, Kao AH, Demirci FY, Kamboh MI, Petri M, Manzi S, Kastner DL, Seldin MF, Gregersen PK, Behrens TW, Criswell LA: *Specificity*

of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS Genet* 2008, 4:e1000084.

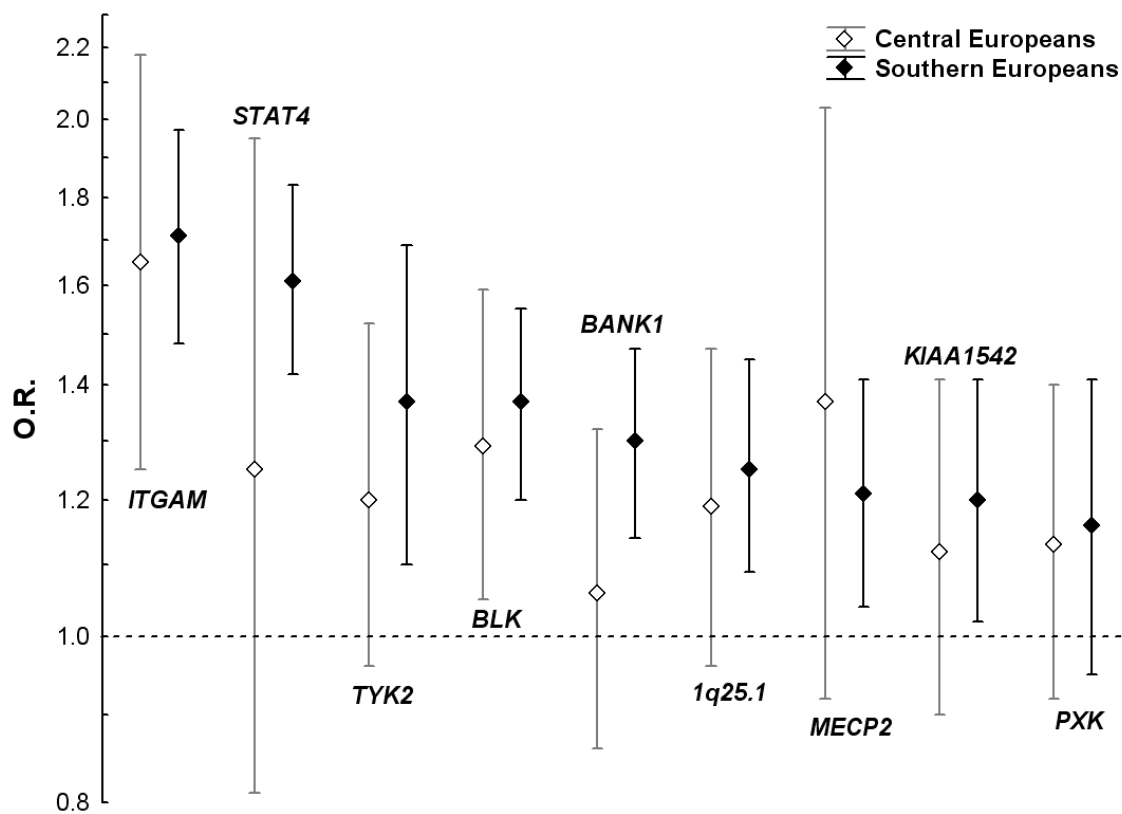
8. Sigurdsson S, Nordmark G, Göring HH, Lindroos K, Wiman AC, Sturfelt G, Jönsen A, Rantapää-Dahlqvist S, Möller B, Kere J, Koskenmies S, Widén E, Eloranta ML, Julkunen H, Kristjansdottir H, Steinsson K, Alm G, Rönnblom L, Syvänen AC: Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 2005, 76:528-537.
9. Cunninghame Graham DS, Akil M, Vyse TJ: Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families. *Rheumatology (Oxford)* 2007, 46:927-930.
10. Sawalha AH, Webb R, Han S, Kelly JA, Kaufman KM, Kimberly RP, Alarcón-Riquelme ME, James JA, Vyse TJ, Gilkeson GS, Choi CB, Scofield RH, Bae SC, Nath SK, Harley JB: Common variants within MECP2 confer risk of systemic lupus erythematosus. *PLoS ONE* 2008, 3:e1727.
11. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Pons-Estel BA, Witte T, D'Alfonso S, Barizzzone N, Danieli MG, Gutierrez C, Suarez A, Junker P, Laustrop H, González-Escribano MF, Martin J, Abderrahim H, Alarcón-Riquelme ME: Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat Genet* 2008, 40:211-216.
12. Cunninghame Graham DS, Vyse TJ, Fortin PR, Montpetit A, Cai YC, Lim S, McKenzie T, Farwell L, Rhodes B, Chad L, Hudson TJ, Sharpe A, Terhorst C, Greenwood CM, Wither J, Rioux JD, CaNIOS GenES Investigators: Association of LY9 in UK and Canadian SLE families. *Genes Immun* 2008, 9:93-102.
13. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, Shifrin N, Petri MA, Kamboh MI, Manzi S, Seldin MF, Gregersen PK, Behrens TW, Ma A, Kwok PY, Criswell LA. *Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus*. *Nat Genet*. 2008, 40(9):1062-4.
14. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, Ortmann W, Kosoy R, Ferreira RC, Nordmark G, Gunnarsson I, Svenungsson E, Padyukov L, Sturfelt G, Jönsen A, Bengtsson AA, Rantapää-Dahlqvist S, Baechler EC, Brown EE, Alarcón GS, Edberg JC, Ramsey-Goldman R, McGwin G Jr, Reveille JD, Vilá LM, Kimberly RP, Manzi S, Petri MA, Lee A, Gregersen PK, Seldin MF, Rönnblom L, Criswell LA, Syvänen AC, Behrens TW, Graham RR. *A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1*

*and IL10 as risk loci for systemic lupus erythematosus.* Nat Genet. 2009, 41:1228-33

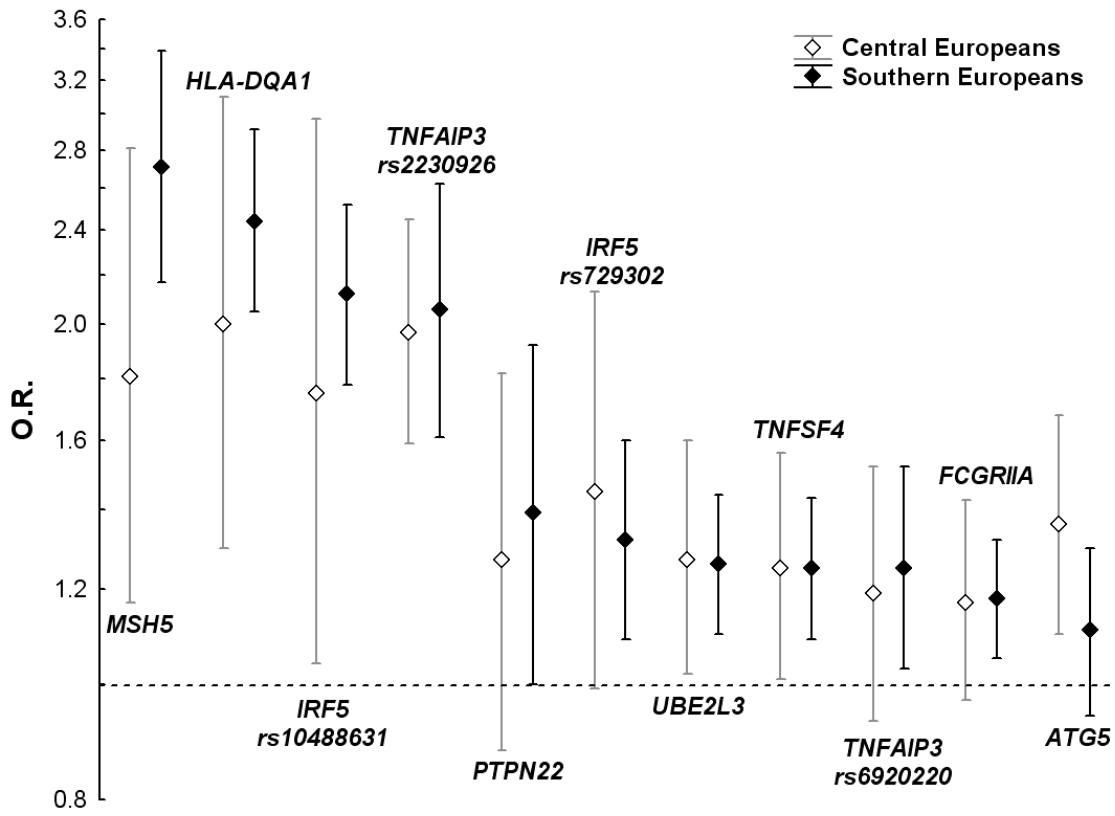
15. Cunninghame Graham DS, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, Moser KL, Rioux JD, Altshuler D, Behrens TW, Vyse TJ. *Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus.* Nat Genet. 2008, 40:83-9.
16. Tian C, Plenge RM, Ransom M, Lee A, Villoslada P, et al. (2008) *Analysis and application of European genetic substructure using 300 K SNP information.* PLoS Genet 4: e4. doi:10.1371/journal.pgen.0040004.
17. Han J, Kraft P, Nan H, Guo Q, Chen C, et al. (2008) *A Genome-Wide Association Study Identifies Novel Alleles Associated with Hair Color and Skin Pigmentation.* PLoS Genet 4: e1000074. doi:10.1371/journal.pgen.1000074.
18. Bersaglieri T, Sabeti PC, Patterson N et al. *Genetic signatures of strong recent positive selection at the lactase gene.* Am J Hum Genet. 2004, 74:1111-20



**Figure S1:** Bias to a stronger association in Southern Europeans than in Central Europeans of 9 SLE loci from Suarez-Gestal et al. [24] Random effect meta-analysis O.R. for the risk allele and their 95 % C.I. are represented in descending order from left to right. *MECP2* data is only from women because this locus is in the X chromosome



**Figure S2:** Comparison of the random effect meta-analysis O.R. for the 12 SNPs associated with SLE in the second phase of our study between the Central Europeans and Southern Europeans. O.R. for the risk allele and their 95 % C.I. are represented in descending order from left to right.



**Estudio de la variabilidad genética de**  
**IRF5 en relación al LES.**

ORIGINAL ARTICLE

# Opposed independent effects and epistasis in the complex association of *IRF5* to SLE

I Ferreira-Neira<sup>1</sup>, M Calaza<sup>1</sup>, E Alonso-Perez<sup>1</sup>, M Marchini<sup>2</sup>, R Scorza<sup>2</sup>, GD Sebastiani<sup>3</sup>, FJ Blanco<sup>4</sup>, I Rego<sup>4</sup>, R Pullmann Jr<sup>5,17</sup>, R Pullmann<sup>5</sup>, CG Kallenberg<sup>6</sup>, M Bijl<sup>6</sup>, FN Skopouli<sup>7</sup>, M Mavromati<sup>7</sup>, S Migliaresi<sup>8</sup>, N Barizzzone<sup>9</sup>, S Ruzickova<sup>10</sup>, C Dostal<sup>10</sup>, RE Schmidt<sup>11</sup>, T Witte<sup>11</sup>, C Papasteriades<sup>12</sup>, I Kappou-Rigatou<sup>12</sup>, E Endreffy<sup>13</sup>, A Kovacs<sup>13</sup>, J Ordi-Ros<sup>14</sup>, E Balada<sup>14</sup>, P Carreira<sup>15</sup>, JJ Gomez-Reino<sup>1,16</sup> and A Gonzalez<sup>1</sup>

<sup>1</sup>Laboratorio Investigacion 2 and Rheumatology Unit, Hospital Clinico Universitario de Santiago, Santiago de Compostela, Spain;

<sup>2</sup>Clinical Immunology, University of Milan and Fondazione IRCCS Ospedale Maggiore Policlinico, Milan, Italy; <sup>3</sup>UO Complessa di Reumatologia, Ospedale S. Camillo – Forlanini, Rome, Italy; <sup>4</sup>Servicio de Reumatologia, CH Universitario Juan Canalejo, A Coruña, Spain; <sup>5</sup>Institute of Clinical Biochemistry, Martin Faculty Hospital, Martin, Slovakia; <sup>6</sup>Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, Groningen, The Netherlands; <sup>7</sup>Pathophysiology Department, Athens University Medical School, Athens, Greece; <sup>8</sup>Rheumatology Unit, Second University of Naples, Caserta, Italy; <sup>9</sup>Department of Medical Sciences and IRCAD, Eastern Piedmont University, Novara, Italy; <sup>10</sup>Molecular Biology and Immunogenetics Department, Institute of Rheumatology, Prague, Czech Republic; <sup>11</sup>Division of Clinical Immunology, Hannover Medical School, Hannover, Germany; <sup>12</sup>Department of Histocompatibility and Immunology, Evangelismos Hospital, Athens, Greece; <sup>13</sup>Paediatrics Department, University of Szeged, Szeged, Hungary; <sup>14</sup>Internal Medicine, Research Laboratory in Autoimmune Diseases, Hospital Vall d'Hebron, Barcelona, Spain; <sup>15</sup>Rheumatology Unit, Hospital 12 de Octubre, Madrid, Spain and <sup>16</sup>Department of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

*Genetic variation in the interferon regulatory factor 5 (IRF5) gene affects systemic lupus erythematosus (SLE) susceptibility. However, association is complex and incompletely defined. We obtained fourteen European sample collections with a total of 1383 SLE patients and 1614 controls to better define the role of the different IRF5 variants. Eleven polymorphisms were studied, including nine tag single nucleotide polymorphisms (SNPs) and two extra functional polymorphisms. Two tag SNPs showed independent and opposed associations: susceptibility (rs10488631,  $P < 10^{-17}$ ) and protection (rs729302,  $P < 10^{-6}$ ). Haplotype analyses showed that the susceptibility haplotype, identified by the minor allele of rs10488631, can be due to epistasis between three IRF5 functional polymorphisms. These polymorphisms determine increased mRNA expression, a splice variant with a different exon 1 and a longer proline-rich region in exon 6. This result is striking as none of the three polymorphisms had an independent effect on their own. Protection was independent of these polymorphisms and seemed to reside in the 5' side of the gene. In conclusion, our results help to understand the role of the IRF5 locus in SLE susceptibility by clearly separating protection from susceptibility as caused by independent polymorphisms. In addition, we have found evidence for epistasis between known functional polymorphisms for the susceptibility effect.*

Genes and Immunity advance online publication, 14 June 2007; doi:10.1038/sj.gene.6364407

**Keywords:** epistasis; haplotype analysis; systemic lupus erythematosus

## Introduction

Systemic lupus erythematosus (SLE (MIM 152700)) is a disease that presents multiple unsolved challenges.<sup>1</sup> Extensive research has identified many immunological abnormalities, but their significance is still unclear. It seems that defects in T- and B-cell tolerance play a critical

role as well as defects in the clearance of apoptotic cells and of immunocomplexes. Recent evidences indicate that there is also an important involvement of the innate immune system involving plasmacytoid dendritic cells (pDCs), toll-like receptors (TLRs) and type I interferons (IFN).<sup>2</sup> In SLE etiology, there is evidence for both an environmental and a genetic component.<sup>3,4</sup> This latter component accounts for the increased risk in siblings, which is about 20 times the risk in the population. However, only a few genetic factors had convincingly been identified. Here, we have explored the impact that genetic variation in the interferon regulatory factor 5 (*IRF5* (MIM 607218)) locus has on SLE susceptibility.

The starting point in this line of research has been a hypothesis-driven genetic study based in the important

Correspondence: Dr A Gonzalez, Laboratorio de Investigacion 2, Hospital Clinico Universitario de Santiago, 15706-Santiago de Compostela, Travesia Choupana, A Coruña, Spain.

E-mail: Antonio.Gonzalez.Martinez.Pedrayo@sergas.es

<sup>17</sup>Current address: Gerontology Research Center, NIA, NIH, Baltimore, MD, USA.

Received 3 April 2007; revised and accepted 15 May 2007

role of type 1 IFN-regulated genes in SLE that found association of three SNPs in the 5' end of *IRF5*.<sup>5</sup> This association, observed in Swedish, Finish and Icelandic subjects,<sup>5</sup> has been extended and confirmed in subsequent studies with Swedish, Spanish, Argentineans, European Americans, British, Danish, German and Korean samples.<sup>6–10</sup> In every case, association has been clear indicating this is one of the major genetic factors in SLE susceptibility. Other aspects of this association have already been revealed, but more genetic investigation is required as association extends over many SNPs in the locus and it is unclear which are relevant. At least two promoters and nine mRNA variants have been described in *IRF5*.<sup>7–9,11</sup> Some of this variation is related with three functional polymorphisms in the genomic sequence. One of them, the rs2004640 SNP, influences the use of three alternative first exons because its T allele creates a donor splice site after exon 1B.<sup>8,9</sup> The rs10954213 SNP determines the length of the 3'-UTR.<sup>7</sup> The A allele at this SNP creates a polyadenylation signal that results in a shorter 3'-UTR and a higher level of expression. Finally, an insertion/deletion polymorphism causes inclusion/loss of ten amino acids in frame that contribute to a proline-rich region of exon 6.<sup>9</sup> Results from some previous genetic studies have indicated that SLE susceptibility was associated to a haplotype containing the donor splice site, T allele of rs2004640, and high RNA levels determined by rs10954213 A allele.<sup>7,8</sup> However, the role of the donor splice site has been questioned on functional grounds<sup>9</sup> and association with the functional polymorphisms fails to explain but a very low fraction of association with the locus.

In our study, some of the previous questions have been solved by identifying two independent effects, susceptibility and protection, that depend on different polymorphisms. In addition, the susceptibility haplotype could be attributed to the epistatic effect of the three functional polymorphisms, each of them without any contribution in isolation.

## Results

### Study characteristics

DNA samples (Table 1) from 14 independent DNA collections from eight European countries were included in the study. Each collection contained approximately 100 SLE patients and 100 ethnically matched healthy controls, all of them Caucasians. A total of 1383 SLE patients and 1614 controls were analyzed. Females predominated both among SLE patients and controls. Significant heterogeneity in control genotypes and problems of Hardy-Weinberg equilibrium (HWE) led us to exclude from analysis samples from the Czech Republic.

A complete coverage of the *IRF5* locus and neighboring sequences was intended with nine tag SNPs polymorphisms selected from HapMap<sup>12</sup> (Figure 1; Supplementary Table 1). One of them was not in HWE in control samples, rs4728142  $P = 0.003$ , and was excluded from further analysis. Information lost was minimal because this SNP was not tagging any other SNP. Remaining SNPs were in HWE. All genotypes were obtained in a single laboratory with a 98.0% success rate obtained uniformly across cases and controls and

**Table 1** DNA samples collected for this study and their female percentage

Collection <sup>a</sup>	DNA samples		Female (%) <sup>b</sup>	
	Control	SLE	Control	SLE
Germany <sup>c</sup>	71	95	39.3	90.4
Czech Republic <sup>d</sup>	95	101	33.7	85.1
Greece 1 <sup>e</sup>	100	100	92.0	91.0
Greece 2 <sup>f</sup>	99	98	67.7	86.7
Hungary <sup>g</sup>	95	94	48.4	90.3
Naples <sup>h</sup>	109	100	100.0	88.9
Milan <sup>i</sup>	108	129	42.5	86.8
Rome <sup>j</sup>	94	84	55.3	92.1
Barcelona <sup>k</sup>	97	93	52.6	92.3
Corunna <sup>l</sup>	94	88	82.9	92.0
Madrid <sup>m</sup>	284	92	68.8	91.3
Santiago <sup>n</sup>	95	109	48.4	96.0
The Netherlands <sup>o</sup>	179	106	59.8	86.8
Slovakia <sup>p</sup>	94	94	93.6	91.5
Total	1614	1383	63.1	89.8

Abbreviation: SLE, systemic lupus erythematosus.

<sup>a</sup>Origin of samples.

<sup>b</sup>There were 33 control and 28 patient samples where the sex was not reported.

<sup>c</sup>Hannover Medical School.

<sup>d</sup>Institute of Rheumatology, Prague.

<sup>e</sup>Athens University Medical School.

<sup>f</sup>Evangelismos Hospital, Athens.

<sup>g</sup>Albert Szent-Györgyi Medical and Pharmaceutical Centre, Szeged.

<sup>h</sup>Second University of Naples.

<sup>i</sup>University of Milan and Fondazione IRCCS Ospedale Maggiore Policlinico.

<sup>j</sup>Ospedale S. Camillo – Forlanini, Rome.

<sup>k</sup>Hospital Val d'Hebron, Barcelona.

<sup>l</sup>CH Universitario Juan Canalejo, Corunna.

<sup>m</sup>Hospital 12 de Octubre, Madrid.

<sup>n</sup>Hospital Clinico Universitario de Santiago.

<sup>o</sup>University Medical Center Groningen.

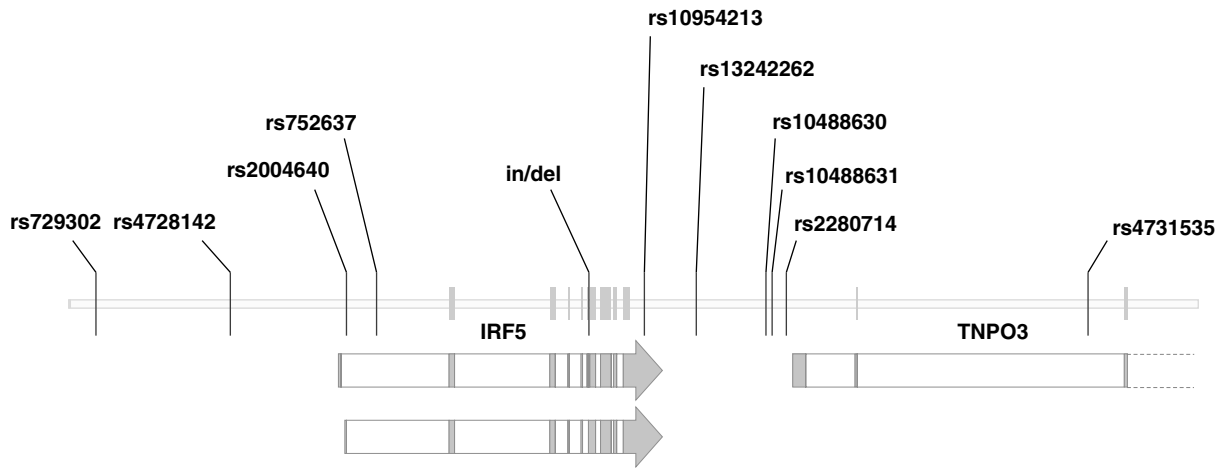
<sup>p</sup>Martin Faculty Hospital, Martin.

different sample collections. Regentyping was used liberally to check for accuracy of results.

Analysis of the linkage disequilibrium (LD) pattern showed high  $D'$  values between the eight tag SNPs indicating that they were in a single LD block (Supplementary Table 2). The  $r^2$  values in all the pair-wise comparisons were below 0.8 (mean value = 0.28; maximum 0.76) as expected for tag SNPs. Therefore it is likely that variation in the *IRF5* locus has been adequately covered and that it was done efficiently with a minimum of redundancy.

### Wide and strong association of *IRF5* with SLE susceptibility

For the joint analysis of differences between cases and controls, we used Mantel-Haenszel analysis. This analysis allows evaluating significance of the differences taking into account that samples come from different collections. Significant differences were observed in all eight tag SNPs (Table 2). They were very clear in three of them, rs729302, rs2004640 and rs10488631. They are, respectively, the most upstream SNP, placed 9 kb before the transcription start site; the SNP creating a splice



**Figure 1** Polymorphisms in the interferon regulatory factor 5 (*IRF5*) locus. The full genomic area covered to select tag SNPs, 42 kb, is included. The CDS for *IRF5* and *TNPO3* are shown on the sequence. Transcripts are shown below. The two transcripts of *IRF5* included in RefSeq are shown.

**Table 2** Allelic frequency differences in the eight *IRF5* tag SNPs between SLE patients and controls

SNP Group	MAF % (n/N)	M-H analysis		B-D test	M-H analysis <sup>a</sup>	
		OR (95% CI)	P		OR (95% CI)	P
rs729302						
SLE	25.1 (588/2340)	0.72 (0.64–0.82)	$2.2 \times 10^{-7}$	0.0028	0.67 (0.59–0.76)	$1.6 \times 10^{-9}$
CRL	31.3 (912/2912)					
rs2004640						
SLE	36.4 (846/2324)	0.73 (0.65–0.81)	$4.3 \times 10^{-8}$	0.2		
CRL	43.7 (1259/2876)					
rs752637						
SLE	29.3 (685/2334)	0.80 (0.71–0.90)	0.00024	0.076		
CRL	34.0 (985/2890)					
rs13242262						
SLE	30.9 (709/2292)	0.81 (0.72–0.91)	0.00047	0.033	0.83 (0.74–0.94)	0.0033
CRL	35.5 (1033/2902)					
rs10488630						
SLE	37.2 (864/2322)	0.89 (0.80–1.00)	0.049	0.032	0.86 (0.76–0.96)	0.0095
CRL	40.3 (1176/2918)					
rs10488631						
SLE	19.0 (437/2294)	2.02 (1.72–2.38)	$7.8 \times 10^{-18}$	0.005	2.17 (1.83–2.57)	$4.8 \times 10^{-20}$
CRL	10.1 (296/2906)					
rs2280714						
SLE	27.1 (628/2314)	0.87 (0.77–0.99)	0.029	0.4		
CRL	29.9 (867/2892)					
rs4731535						
SLE	49.3 (1145/2320)	1.12 (1.00–1.25)	0.049	0.0013	1.18 (1.05–1.33)	0.0044
CRL	46.3 (1348/2908)					

Abbreviations: *IRF5*, interferon regulatory factor 5; SNPs, single nucleotide polymorphisms; SLE, systemic lupus erythematosus; MAF, minor allele frequencies; M-H, Mantel-Haenszel; B-D, Breslow-Day.

<sup>a</sup>Analysis repeated after excluding the sample collection causing significant effect heterogeneity: Milan in rs729302 and Germany in the other SNPs.

donor site; and a SNP that is about 4 kb downstream the gene. The rs10488631 SNP showed, by much, the most extreme difference between cases and controls (G allele frequencies 19.0 and 10.1%, respectively). Given the large sample size of the joint analysis, the *P*-value for this difference was very low ( $P < 10^{-17}$ ). Marked frequency differences between SLE patients and controls extended for about 25 kb. This together with the low  $r^2$  between the involved tag SNPs indicated that the *IRF5* locus

likely contains several polymorphisms modifying SLE susceptibility.

We checked for effect size heterogeneity between the sample collections using the odds ratio (OR) obtained with allelic frequencies in SLE cases and controls as effect size and the Breslow-Day as heterogeneity test (Table 2). There was significant heterogeneity in OR values for five SNPs. In each of the five SNPs, heterogeneity was caused by divergent results in a single sample collection. For the

rs729302 SNP, the collection from Milan showed an increased frequency of the G allele in SLE patients in relation with the controls from the same town. For the other four SNPs, rs10488630, rs10488631, rs4731535 and rs13242262, heterogeneity was due to the samples of the German collection. Once data from the discordant collection for each of these SNPs were excluded from analysis, significant heterogeneity disappeared and OR estimates became more precise (Table 2), but interpretation of results did not change. We looked for possible causes of the discordant results by reviewing the accuracy of genotypes and the available clinical data from the two-outlier sample collections, but no likely cause was found. In subsequent analyses, we have considered whether the heterogeneity introduced by German and Milanese samples modifies the interpretation of results, and we have signaled this circumstance in each case.

Checking different genetic models by logistic regression showed that additive models accounted adequately for the effect of all tag SNPs, and they were significantly better than either dominant or recessive models in many of them (Supplementary Table 3). This was especially clear in the most strongly associated SNPs, rs10488631, rs2004640 and rs729302, where only the additive model fitted genotype effects.

#### *Genotype association conditional in the other tag SNPs*

Given the broad association appreciated, it became necessary to analyze the effects of each tag SNP conditional in other SNPs to distinguish between independent effects and those due to LD. Conditional logistic regression with additive models showed that only two SNPs, rs2004640 and rs729302, retained significant association when evaluated conditional on the most strongly associated SNP, rs10488631 (Figure 2a). Therefore, the effect of these two SNPs could have an independent contribution to SLE susceptibility. On the other side, association of rs10488631 with SLE was almost unaffected by accounting for any of the other SNPs, as shown in Figure 2b for the rs10488631 OR conditional on rs2004640, on rs729302 or on both SNPs. These results indicated that the large effect of the rs10488631 SNP was almost completely independent of any of the other tag SNPs.

Further analysis of the rs2004640 and rs729302 SNPs showed that none of them remained significant when considered conditional on the other SNP together with rs10488631 (Figure 2b). In consequence, only one of these two SNPs contributed independently to association, but it was unclear which of them as inclusion of one cancelled the effect of the other. Excluding samples from Germany and Milan modified this result: in this case, the rs729302 SNP remained significantly different whereas the rs2004640 SNP did not (Supplementary Figure 1).

#### *Haplotype analysis of tag SNPs*

We performed haplotype analysis with the eight *IRF5* tag SNPs. There were only six haplotypes with a frequency over 5% (Table 3). They represented 83.4 and 82.4% of the chromosomes in SLE patients and controls, respectively. Frequency distribution of haplotypes was extremely different in SLE patients and controls ( $P = 2.7 \times 10^{-14}$ ). It showed readily two opposing effects: a susceptibility haplotype, #6, and two protective haplotypes, #1 and 2

(Figure 3). The most different discordant frequency was that of the susceptibility haplotype. It was identified by the minor allele, G, of the rs10488631 SNP. Haplotype #6 contained the common alleles at all the other SNPs except rs4731535. The common alleles were present in several other haplotypes and the minor allele of rs4731535 was included in two other haplotypes, a neutral, #3, and a protective haplotype, #2. A haplotype that has previously been reported as conferring SLE susceptibility, defined by the T allele of rs2004640 and the A allele of rs2280714,<sup>8</sup> was split in three haplotypes, #4–6, in our analysis and only haplotype #6 showed association.

Two haplotypes were significantly less frequent in SLE patients than in controls, #1 and #2 (Figure 3). These differences were not due to an effect of balancing out the susceptibility effect of haplotype #6 because the two protective haplotypes together represented only about 20% of the total. The protective haplotypes were uniquely characterized by containing the minor allele, C, of the rs729302 SNP. All the other alleles in these two haplotypes were also present in neutral haplotypes. Especially interesting was the comparison of haplotypes #2 and #3 because they only differed in the allele of rs729302. In haplotype #2, which was protective, the rs729302 allele was C, whereas in the neutral haplotype #3 it was A. We explored more carefully the contribution of SNPs rs729302 and rs2004640 to the protective haplotypes. For this analysis, only three SNPs, rs729302, rs2004640 and rs10488631, were included in a new estimation of haplotype frequencies (Supplementary Table 4b). The table of haplotype frequencies showed that only the rare allele of SNP rs729302 correlated unambiguously with the protective haplotype.

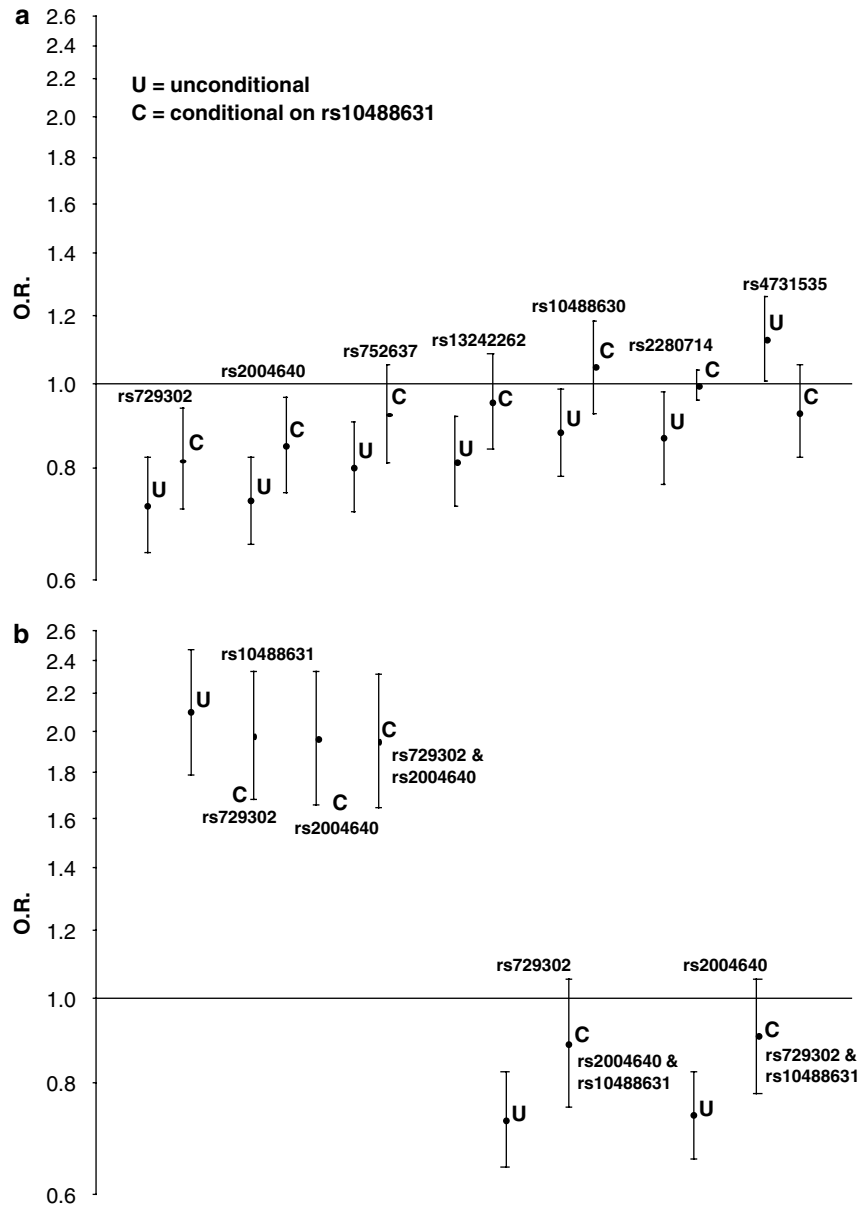
In summary, association of the *IRF5* locus was due to two independent and opposing effects that depended on haplotypes defined by the minor alleles of two SNPs, rs10488631 and rs729302.

#### *Two additional functional polymorphisms in IRF5*

Two new functional polymorphisms have been described in *IRF5* during progress of our study, an insertion/deletion polymorphism and the rs10954213 SNP.<sup>7,9</sup> We have genotyped them in 486 control and 459 SLE samples from Slovakia, Rome, Naples, Hungary and Santiago with a 98.3% success rate and without significant deviation from HWE. The two were already well covered by the tag SNPs of our panel. SNP rs10954213 had  $r^2$  of 0.93 with SNP rs13242262, and the insertion/deletion polymorphism had  $r^2$  of 0.90 with SNP rs4731535. Therefore, results were very similar to the observed for these two tag SNPs.

The rs10954213 SNP was moderately different in SLE patients and controls (Table 4). The insertion/deletion polymorphism was not significantly different. When data from rs10954213 were considered conditional on the rs10488631 SNP, the association disappeared. The unconditional OR of this SNP, 0.82 (0.67–0.99), changed to a conditional OR of 0.94 (0.76–1.16). These results indicate that none of these two new functional polymorphisms have a significant contribution to SLE susceptibility by themselves.

Haplotype analysis considering all ten polymorphisms showed that the two new polymorphisms were irrelevant for the distribution of haplotype frequencies



**Figure 2** Association of the tag SNPs in *IRF5* with SLE susceptibility conditional in other SNPs. (a) Odds ratios of each of the tag SNPs unconditional (U) or conditional (C) on the genotypes of the rs10488631 SNP according to an additive logistic regression model. SNP order follows the sequence. (b) Analysis of the three SNPs with independent contribution to the association in A. Their unconditional (U) and conditional (C) OR on the specified SNP or combination of SNPs are shown. Ranges represent the 95% C I of the OR. Logarithmic scale has been used for the y axis; the x axis is placed at the identity level OR = 1.

(Supplementary Table 5a). To focus on the three functional polymorphisms, we obtained the frequency distribution of haplotypes considering only these polymorphisms and the rs729302 and rs10488631 SNPs (Figure 4 and Table 5). There were six haplotypes with frequencies over 5% that accounted for 87.7 and 91.7% of all haplotypes in cases and controls, respectively. The effects of the individual SNPs were as described in the previous analysis: protection correlated with the minor allele of rs729302 and susceptibility with the minor allele of rs10488631. None of the putative functional polymorphisms correlated with any of the two effects, susceptibility or protection. However, there were evidences of an epistatic effect between them. The susceptibility haplotype, #6, included a combination of alleles

for the three functional SNPs that was not present in any other common haplotype (even in haplotypes with frequencies over 2%). This combination, 'T-in-A', included the allele inducing alternative splicing (T of rs2004640), the insertion prolonging the proline-rich region of exon 6 and the short 3'-UTR associated with higher mRNA expression levels (A of rs10954213). None of the other common combinations of these three polymorphisms was associated with an increase in disease susceptibility. On the other side, the alleles corresponding to the three functional polymorphisms were irrelevant in the protective haplotypes.

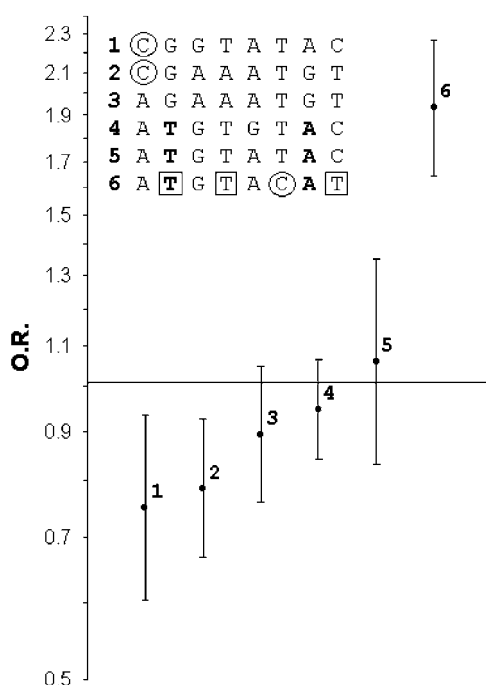
The rs2004640 SNP and the two SNPs (rs4731535 and rs13242262) that were tag SNPs for the other two functional polymorphisms showed also a unique combi-



**Table 3** Haplotype frequencies of the tag SNPs in *IRF5* in SLE patients and controls

No.	Haplotype	% in controls (n)	% in SLE (n)	OR (95% CI)
1	CGGTATAC	7.7 (225)	5.9 (138)	0.75 (0.6–0.9)
2	CGAAATGT	14.3 (417)	11.6 (271)	0.79 (0.7–0.9)
3	AGAAATGT	14.0 (410)	12.7 (298)	0.89 (0.8–1.1)
4	ATGTGTAC	31.3 (915)	30.1 (706)	0.95 (0.8–1.1)
5	ATGTATAC	5.2 (150)	5.4 (127)	1.06 (0.8–1.4)
6	ATGTACAT	10.0 (293)	17.7 (415)	1.93 (1.6–2.3)

Abbreviations: SNPs, single nucleotide polymorphisms; *IRF5*, interferon regulatory factor 5; SLE, systemic lupus erythematosus. Only haplotypes with frequency over 5% are shown. Haplotypes are ordered according to their relative frequencies in cases and controls. SNP order in the haplotypes are as in Figure 3.



**Figure 3** Haplotype analysis of the tag SNPs in *IRF5*. OR and their 95% CI are represented for each of the six haplotypes with frequency over 5%. Alleles included in each haplotype are signalled following order in the sequence: rs729302, rs2004640, rs752637, rs13242262, rs10488630, rs10488631, rs2280714 and rs4731535. Alleles identifying susceptible and protective haplotypes are encircled. Alleles previously reported as forming a susceptible haplotype are shown in bold.\* Alleles of the rs2004640 SNP and of two tag SNPs (inside rectangles) that are in tight LD with two other functional SNPs form a unique combination in haplotype #6.

nation of alleles in the susceptibility haplotype #6 also when all the sample collections were included (Figure 3, Supplementary Table 4A). Therefore, results implicating an epistatic effect between the three functional polymorphisms were consistently observed.

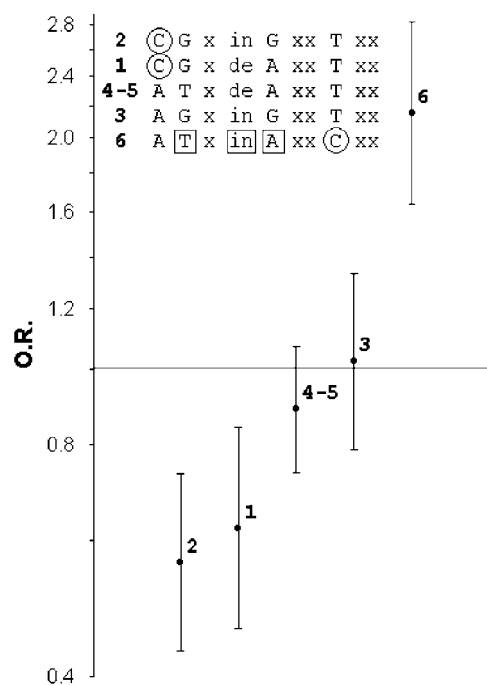
Haplotype analysis restricted to females also showed the results that have been described above for the unstratified set of samples (Supplementary Tables 6A and B). We performed a correlation analysis between the

**Table 4** Allelic frequencies of the two new *IRF5* functional SNPs

SNP Group	MAF % (n/N)	M-H analysis		B-D test
		OR (95% CI)	P	
in/del				
SLE	50.8 (438/862)	1.10 (0.91–1.32)	0.3 <sup>a</sup>	0.3
CRL	46.9 (447/954)			
rs10954213				
SLE	30.8 (260/844)	0.78 (0.64–0.95)	0.015	0.5
CRL	36.4 (334/918)			

Abbreviations: *IRF5*, interferon regulatory factor 5; SNPs, single nucleotide polymorphisms; MAF, minor allele frequencies; M-H, Mantel-Haenszel; B-D, Breslow-Day; in/del, insertion/deletion; SLE, systemic lupus erythematosus.

<sup>a</sup>Given the lower number of samples, two of the tag SNPs were also not significantly different in this set of samples, rs10488630 and rs4731535.



**Figure 4** Haplotype analysis restricted to the three functional polymorphisms and the two SNPs identifying SLE associated haplotypes. OR and their 95% CI are represented for the five haplotypes with frequency over 5% conserving the haplotype numbers given in Figure 3. Alleles included in each haplotype are in the sequence order maintaining an x in the place of tag SNPs not included in the estimation of these haplotypes: rs729302, rs2004640, x, insertion/deletion, rs10954213, x, x, rs10488631, x, x. Alleles identifying susceptible and protective haplotypes are encircled. Alleles of the functional SNPs form a unique combination in haplotype #6 (included in rectangles).

rs10488631 and rs739302 genotypes and the 1997 ACR classification criteria for SLE, but no correlation was found. We also checked that inclusion of the rs4728142 SNP or of Czech Republic sample collection data, which were excluded from analysis as mentioned above, did not significantly modify the results.

**Table 5** Haplotype frequencies of three functional polymorphisms and the two SNPs identifying SLE associated haplotypes

No.	Haplotype	% in controls (n)	% in SLE (n)	OR (95% CI)
2	CGx in GxxTxx	17.8 (170)	10.8 (100)	0.56 (0.4–0.7)
1	CGx de AxxTxx	12.8 (122)	8.4 (77)	0.62 (0.5–0.8)
4–5	ATx de AxxTxx	38.9 (372)	36.1 (333)	0.89 (0.7–1.1)
3	AGx in GxxTxx	13.7 (131)	14.0 (129)	1.02 (0.8–1.3)
6	ATx in AxxCxx	9.5 (91)	18.4 (170)	2.15 (1.6–2.8)

Abbreviations: SNPs, single nucleotide polymorphisms; SLE, systemic lupus erythematosus.

Only haplotypes with frequency over 5% are shown. Haplotypes are ordered according to Table 3. Allele notation and order in the haplotypes are as in Figure 4.

## Discussion

Our results have confirmed the striking association of genetic variation in the *IRF5* locus with SLE susceptibility.<sup>5–10</sup> Association is already very clear with *P*-values that make the results of an unprecedented strength in SLE genetics. It has been found in several populations of European Caucasians, in Argentineans, in Koreans and in a small number of Indo-Pakistani families. It is still unclear whether all ethnic groups will be affected similarly. We have included a diverse array of European Caucasian sample collections, and although there was some heterogeneity between them, it did not follow any identifiable pattern. Our results have shown the most extreme differences to date, probably due to the choice of tag SNPs, to the large number of samples and the use of a case-control design. However the SLE-associated haplotypes are present only in a fraction of SLE patients and it is unclear if they are from a specific subset of patients. Previous studies had already signaled, and we have confirmed this, that there are not specific SLE ACR classification criteria related with genetic variability in the *IRF5* locus.<sup>5,7,8</sup> However, data on disease activity and accumulated damage were not included in any of these analyses, and these SLE features are better candidates to be related with *IRF5* variation because they have shown correlation with the overexpression of type 1 IFN response genes.<sup>13,14</sup>

Our analysis has contributed to clarify the contribution of *IRF5* to SLE in several ways. First, it has allowed the distinction of two independent and opposed effects in the locus. Magnitude of the susceptibility effect was larger than the protective one: the susceptible haplotype was about twice more frequent in SLE patients than in controls, whereas the protective haplotypes were near 1.5-fold less frequent. This is the first time the two opposed effects have been shown in a clear way and that they have been shown to be independent. Only a previous report mentioned a protective haplotype,<sup>7</sup> but it was unclear whether the protective effect was related with the minor allele, G, of the rs2004640 SNP and, therefore, a consequence of the susceptibility effect associated with its major T allele. However, the G of rs2004640 in our study was present in three haplotypes and only two of them were significantly less frequent in SLE patients than in controls indicating that it could not

account for SLE protection. In the other reports,<sup>5,6,8–10</sup> an independent protective effect was not observed because they included less SNPs and their haplotype analysis lacked sensitivity.

Susceptibility was accounted for by the minor allele of one of the tag SNPs, rs10488631, which has not identifiable functional relevance. This SNP has already been studied in UK families without being remarked as special but only about 30% of the families were informative for this SNP and this, likely, hampered analysis.<sup>7</sup> The SNP has a high  $r^2$  with ten other SNPs according to HapMap data (Supplementary Figure 2) that are, all of them, in the *TNPO3* gene. As the minor allele of rs10488631 identifies unambiguously the susceptibility haplotype, it is clear that this haplotype extends into the *TNPO3* gene. This raises the possibility that variation in this large gene, coding for a protein involved in transport to the nucleus of some members of the spliceosome, could play a role in SLE susceptibility. However, it seems unwarranted to relate this gene to SLE in the absence of any other evidence and because none of the SNPs showing high values of  $r^2$  with rs10488631 was predicted to have functional relevance by bioinformatics analysis. A better-grounded alternative is to consider rs10488631 as a tag SNP for the haplotype formed by the 'T-in-A' alleles of the three identified functional *IRF5* polymorphisms. This haplotype includes the donor splice site for exon 1B, the two extra repetitions in exon 6 that extend a proline-rich region and the short 3'-UTR that determines increased mRNA expression of *IRF5*. What is most notable is that none of these three alleles seem able to affect SLE susceptibility by themselves or even in the combinations that are observed in common haplotypes. Results pointing to a similar conclusion have also been reported recently.<sup>15</sup> It is unclear whether the three are required for increased SLE risk, it could be that 'T-in-x' or 'x-in-A' haplotypes will be enough as they are also restricted to the susceptibility haplotype #6. The latter haplotype, 'x-in-A' seems to be favored by the experiments that show that the donor splice SNP rs2004640 has little consequence in the pattern of expressed *IRF5* variants.<sup>9</sup> The main piece of evidence required to complete the hypothesis of epistasis is to define the role of the insertion/deletion polymorphism. The effects of the donor splice site and the 3'-UTR variants have already been sufficiently demonstrated.<sup>7–9</sup> On the contrary, the effect of the different length of the proline-rich region has not been explored. This type of region has been involved in protein–protein interactions.<sup>16</sup> It will also be necessary to find why *IRF5* function is dramatically different when the functional alleles are in the same haplotype. In this respect, it has already been reported that each different *IRF5* mRNA variant induces expression of a different pattern of IFN subtypes in virus-infected cell lines.<sup>11</sup>

Regarding the protective effect, the rs729302 SNP that identifies the protective haplotypes has not any other SNP in tight LD according to HapMap. It is located about 9 kb upstream the *IRF5* gene and has no obvious consequence. Given this position, it is tempting to speculate that it could be in LD with unknown SNPs that affect *IRF5* promoters.<sup>11</sup>

We think important to remark the significant contribution that HapMap-selected tag SNPs have had to our results. Coverage of variation in the *IRF5* locus has been

very good as evidenced not only in the LD analysis between the tag SNPs, but also by the two polymorphisms included in a second part of the analyses that were already well covered. This efficient coverage allowed us to explore the full locus with a small number of SNPs and permitted our new findings. Only, another report has used a comparable coverage and this is the report finding a more complete picture of the effects determined by *IRF5* variation before ours.<sup>7</sup> We have obtained some additional details because we used a case-control design with large sample size that is more powerful than a family association study; and because, thanks to the work of others, we were aware of the three functional polymorphisms in *IRF5*.

Although *IRF5* is likely to have an important participation in SLE autoimmunity, most studies of this transcription factor have been related with response to viral infections and with its involvement in regulation of cell cycle and survival of cancer cells.<sup>17–20</sup> *IRF5* is a critical mediator of signals triggered by TLRs.<sup>21–23</sup> Its phosphorylation and nuclear translocation leads to the overexpression of many proinflammatory genes including cytokines, chemokines and type 1 IFN.<sup>22,24,25</sup> The specific link of *IRF5* to SLE susceptibility could reside in the antigen–antibody complexes containing DNA and RNA.<sup>1,2</sup> These complexes are taken by pDCs and interact with TLR7/8 and 9 leading to activation of DCs and of autoreactive B cells.

In conclusion, our data indicate that *IRF5* is a major factor in SLE genetics, with variants playing both a susceptible and a protective role. Likely candidate for the susceptible effect is an epistatic haplotype determining increased mRNA expression of an *IRF5* isoform including exon 1B and an extended proline-rich region, although it is unclear whether the three alleles are necessary. If this hypothesis is confirmed by other studies, it will be a very clear example of epistasis as none of the three *IRF5* variants individually, and none of the other available haplotypes, has significant effect. The protective effect seems independent of any known functional polymorphism. Future research should address the mechanisms behind epistasis and the protective effect.

## Materials and methods

### Sample collection

We looked for DNA samples from about 100 SLE patients and 100 ethnically matched healthy controls from different European countries (Table 1). Most of these samples have been already described;<sup>26,27</sup> new collections were from Corunna in Spain, Martin in Slovakia and Groningen in the Netherlands. Significant heterogeneity and lack of HWE in several SNPs among controls led to the exclusion from further study of samples from the Czech Republic. Initial heterogeneity of genotypes of controls from Madrid disappeared after the inclusion of more controls. Patients and controls from the German collection have been included also in the study done by Kozzyrev *et al.*<sup>9</sup> A total of 1383 SLE and 1614 control DNA samples were obtained as detailed in Table 1. All SLE patients met the revised American College of Rheumatology classification criteria.<sup>28</sup> Patients and controls provided written informed consent. Sample collection

was approved by the respective ethical committees and the study was reviewed and approved by the Comité Ético de Investigación Clínica de Galicia (Spain).

### Genotyping

We have studied 11 polymorphisms in the *IRF5* locus (Figure 1). Initially, we selected nine tag SNPs that covered variability in the *IRF5* gene and neighboring sequences (from 10 kb upstream to 20 kb downstream) based on HapMap CEU data (Supplementary Table 1).<sup>12</sup> For this selection, we fixed SNPs that have been shown to be associated with SLE susceptibility and used the pair-wise option of the Haploview software<sup>29</sup> with a threshold of  $r^2 = 0.8$  and a minor allele frequency of 10%. These nine SNPs were amplified in a single PCR reaction done with the Qiagen Multiplex PCR kit (Qiagen, CA, USA) on 30 ng of genomic DNA. PCR products were purified by Exo-SAP digestion with Exonuclease I (Epicentre, Madison, WI, USA) and Shrimp Alkaline Phosphatase (Amersham Biosciences, Barcelona, Spain). Single-base extension reactions with the SNaPshot Multiplex Kit (Applied Biosystems, Foster City, CA, USA) were done. Samples were analyzed in the Abi Prism 3130xl Genetic Analyzer (Applied Biosystems).

Two functional polymorphisms in *IRF5* that were described during this study were included.<sup>7,9</sup> The rs10954213 SNP was genotyped with a fluorogenic 5' nuclease assay (TaqMan MBG Genotyping Assay, Applied Biosystems) performed in a total volume of 10  $\mu$ l containing 25 ng of genomic DNA, 1  $\times$  TaqMan Universal PCR Master Mix, No Amperase UNG (Applied Biosystems) and 1  $\times$  Assay Mix containing primers and labeled probes. A Chromo4 real-time PCR system (MJ Research, Waltham, MA, USA) was used to run this assay. An insertion/deletion polymorphism of 30 bp in exon 6 has been genotyped by sizing the amplicons resulting from conventional PCR in 3% agarose gels. Sequences of primers and probes are available in Supplementary Table 1.

### Statistical analysis

Analysis of results was based on the following programs: Haploview;<sup>29</sup> Arlequin (10<sup>6</sup> permutations used thoroughly);<sup>30</sup> Statistica 7.0 (StatSoft, Tulsa, OK, USA) and Phase 2.1.<sup>31</sup> HWE was tested in control samples with a threshold of 0.05 without correction for multiple testing. The exact test of population differentiation<sup>32</sup> implemented in Arlequin was used to explore differences in unphased genotypes between the control samples from each collection. Only controls were included because the objective of this analysis is to explore if the sampled populations can be considered in relation to *IRF5* variation either as a single one or as multiple different populations. To account for possible variability in effect size between collections, case-control allele frequencies were compared with the Mantel–Haenszel test for the 2  $\times$  2 contingency tables stratifying by collection of samples. In other words, response variable has been disease status; explanatory variable, allele frequency and control variable, the sample collection. Homogeneity of effect size across collections was assessed with the Breslow–Day test using as effect size the OR obtained with the allelic frequencies in SLE patients and controls. Likelihood ratio tests for the additive, dominant and recessive genetic models were

obtained relative to the codominant model. Likelihoods for the fit of each model were calculated with univariate logistic regression. Multivariate logistic regression analysis was used to evaluate the effect of each SNP conditional on the remaining. A simplified additive model without interaction parameters was used for this analysis with genotype codes 0, 1 and 2 for genotypes AA, Aa and aa, respectively. Haplotypes were estimated with the Phase 2.1 software (default parameters) that implements a Bayesian algorithm. Further analysis was done by supervised removal of SNPs that did not discriminate between individually associated haplotypes. OR for each haplotype and its 95% confidence intervals (CI) were calculated taking as reference all chromosomes not bearing the haplotype. Assessment of the possible functional relevance of the SNPs was done with the PUPAsuite that includes several prediction software applications (<http://pupasuite.bioinfo.cipf.es/>).

## Acknowledgements

Cristina Fernandez-Lopez and Marta Picado have provided outstanding technical assistance. This work has been supported by Fondo de Investigacion Sanitaria, Instituto de Salud Carlos III (Spain), Grants 04/1651 and 06/0620 that are partially financed by the FEDER program of the EU and by Grants from the Xunta de Galicia. SR and CD were supported by Grant 00023728 of the Ministry of Health of the Czech Republic. RES and TW were supported by BMBF, KN Rheuma C2.12. Work by SD'A was supported by Telethon (Grant E1221) and the CARIPLO Foundation.

## Conflict of interests

The authors have no conflict of interests to declare in relation with this manuscript.

## References

- 1 Croker JA, Kimberly RP. SLE: challenges and candidates in human disease. *Trends Immunol* 2005; **26**: 580–586.
- 2 Pascual V, Farkas L, Banchereau J. Systemic lupus erythematosus: all roads lead to type I interferons. *Curr Opin Immunol* 2006; **18**: 676–682.
- 3 Nath SK, Kilpatrick J, Harley JB. Genetics of human systemic lupus erythematosus: the emerging picture. *Curr Opin Immunol* 2004; **16**: 794–800.
- 4 Croker JA, Kimberly RP. Genetics of susceptibility and severity in systemic lupus erythematosus. *Curr Opin Rheumatol* 2005; **17**: 529–537.
- 5 Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 2005; **76**: 528–537.
- 6 Demirci FY, Manzi S, Ramsey-Goldman R, Minster RL, Kenney M, Shaw PS et al. Association of a common interferon regulatory factor 5 (IRF5) variant with increased risk of systemic lupus erythematosus (SLE). *Ann Hum Genet* 2006; **71**: 308–311.
- 7 Graham DS, Manku H, Wagner S, Reid J, Timms K, Gutin A et al. Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. *Hum Mol Genet* 2007; **16**: 579–591.
- 8 Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat Genet* 2006; **38**: 550–555.
- 9 Kozyrev SV, Lewen S, Reddy PM, Pons-Estel B, Witte T, Junker P et al. Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus. *Arthritis Rheum* 2007; **56**: 1234–1241.
- 10 Shin HD, Sung YK, Choi CB, Lee SO, Lee HW, Bae SC et al. Replication of genetic effects of interferon regulatory factor 5 (IRF5) on systemic lupus erythematosus in a Korean population. *Arthritis Res Ther* 2007; **9**: R32.
- 11 Mancl ME, Hu G, Sangster-Guity N, Olshalsky SL, Hoops K, Fitzgerald-Bocarsly P et al. Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J Biol Chem* 2005; **280**: 21078–21090.
- 12 Consortium IH. A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 13 Baechler EC, Batliwalla FM, Karypis G, Gaffney PM, Ortmann WA, Espe KJ et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 2003; **100**: 2610–2615.
- 14 Bennett L, Palucka AK, Arce E, Cantrell V, Borvak J, Banchereau J et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 2003; **197**: 711–723.
- 15 Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci USA* 2007; **104**: 6758–6763.
- 16 Li SS. Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem J* 2005; **390**: 641–653.
- 17 Barnes BJ, Moore PA, Pitha PM. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *J Biol Chem* 2001; **276**: 23382–23390.
- 18 Mori T, Anazawa Y, Iizumi M, Fukuda S, Nakamura Y, Arakawa H et al. Identification of the interferon regulatory factor 5 gene (IRF-5) as a direct target for p53. *Oncogene* 2002; **21**: 2914–2918.
- 19 Hu G, Mancl ME, Barnes BJ. Signaling through IFN regulatory factor-5 sensitizes p53-deficient tumors to DNA damage-induced apoptosis and cell death. *Cancer Res* 2005; **65**: 7403–7412.
- 20 Yanai H, Chen HM, Inuzuka T, Kondo S, Mak TW, Takaoka A et al. Role of IFN regulatory factor 5 transcription factor in antiviral immunity and tumor suppression. *Proc Natl Acad Sci USA* 2007; **104**: 3402–3407.
- 21 Schoenemeyer A, Barnes BJ, Mancl ME, Latz E, Goutagny N, Pitha PM et al. The interferon regulatory factor, IRF5, is a central mediator of toll-like receptor 7 signaling. *J Biol Chem* 2005; **280**: 17005–17012.
- 22 Takaoka A, Yanai H, Kondo S, Duncan G, Negishi H, Mizutani T et al. Integral role of IRF-5 in the gene induction programme activated by toll-like receptors. *Nature* 2005; **434**: 243–249.
- 23 Ouyang X, Negishi H, Takeda R, Fujita Y, Taniguchi T, Honda K et al. Cooperation between MyD88 and TRIF pathways in TLR synergy via IRF5 activation. *Biochem Biophys Res Commun* 2007; **354**: 1045–1051.
- 24 Barnes BJ, Kellum MJ, Field AE, Pitha PM. Multiple regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes. *Mol Cell Biol* 2002; **22**: 5721–5740.

- 25 Barnes BJ, Richards J, Mancl M, Hanash S, Beretta L, Pitha PM *et al*. Global and distinct targets of IRF-5 and IRF-7 during innate response to viral infection. *J Biol Chem* 2004; **279**: 45194–45207.
- 26 Ferreiros-Vidal I, Gomez-Reino JJ, Barros F, Carracedo A, Carreira P, Gonzalez-Escribano F *et al*. Association of PDCD1 with susceptibility to systemic lupus erythematosus: evidence of population-specific effects. *Arthritis Rheum* 2004; **50**: 2590–2597.
- 27 Ferreiros-Vidal I, D'Alfonso S, Papasteriades C, Skopouli FN, Marchini M, Scorza R *et al*. Bias in association studies of systemic lupus erythematosus susceptibility due to geographical variation in the frequency of a programmed cell death 1 polymorphism across Europe. *Genes Immun* 2007; **8**: 138–146.
- 28 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**: 1725.
- 29 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 30 Schneider S, Roesli D, Excoffier L. *Arlequin ver 2000 A: Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva: Switzerland, 2000.
- 31 Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
- 32 Goudet J, Raymond M, de Meeus T, Rousset F. Testing differentiation in diploid populations. *Genetics* 1996; **144**: 1933–1940.

Supplementary information accompanies the paper on Genes and Immunity web site (<http://www.nature.com/gene>)

## Supplementary table 1

### Oligonucleotides used in the study of the IRF5 locus

SNP ID	position Chr 7 build 126	distance from TSS*	Minor Allele	PCR primers #		Minisequencing primers	length of minisequence	Alleles in minisequence	Alleles as in plus strand
				forward	reverse				
rs729302	128356196	-9034	G	acccactctctgcacatctg	ccccaaccaagggtgaacc	ctgggtggactctggtgtgtaggtgatcctg	31	T>G	A>C
rs4728142	128361203	-4027	T	cgggcagcttcctccacagg	agccgagattgcaccattgc	cacagctgggccttcctccccatttcttactaacac \$	61	C>T	G>A
rs2004640 &	128365537	307	C	cgctactgcccccaagtcta	gagggagagcagcagagc	gccggccgggcgcaccctgctgtaggcacc \$	56	A>C	T>G
rs752637	128366656	1426	T	tgccacacctggtccagacag	ggaggccctctggcaagacc	accctgaccctgggagggaagc	21	C>T	G>A
rs13242262	128378600	13370	A	gatgtcacatatctgggtttccagc	tggggaagtcagccactcg	tggcaaatatgaaaacactgtcatacattcctcaaaggcaa	41	T>A	T>A
rs10488630	128381184	15954	G	tgataacggtttcatggatgt	tcatacgcgttacattttaccc	ggtcaggaaactgtctacttgggtggccatagtagaaaaacc \$	67	A>G	A>G
rs10488631	128381419	16189	G	tgtaagcgtgatgagctttgg	cagcccttcacagcoatcag	tgtttctttgtttttactttcattcactgcctttagctcgaaatgggttc	51	A>G	T>C
rs2280714	128381961	16731	G	caaggaggagtaagcaaggaac	cccagttttggctcttctc	ggtcacattggtgatgctc	19	A>G	A>G
rs4731535	128393203	27973	C	attctggcacacgctacaac	tgttggtgcacaaccaatctca	gtatctagagtagtcagactcaaacattaagtacgaagttaattgc	46	T>C	C>T
TaqMan probes									
rs2004640 €				proprietary information of Appliedbiosystems					
rs10954213	128376663	11433	G	gtccaagaacctggagcagaaata:ggagcaactggtagtagtaaaagctctgtttttaacatt[c/t]attaat					
in/del exon 6	128374588	9358	In	gaggtggcactgacagccgctc	ggggaccagcaccacg				

#### Notes:

# PCR primers were selected with Primer 3 and FastPCR softwares. They were checked to avoid formation of primer dimers between the 18 primers included in the reaction

\* TSS of IRF5 RefSeq isoform a

\$ These oligonucleotides were extended with a 25 nt 5' tail that has not homology with human sequences: gggtctgttcttcgttgacatgagg

& The sequence surrounding this SNP has a high GC content and many primer designs failed

€ Assay used to validate genotypes obtained with minisequencing

## Supplementary table 2

Pair wise comparison of the genotypes at 8 SNPs in IRF5 between European collections of control samples

	Germany	Barcel.	Czech R.	Corunna	Slovakia	Greece	Nether.	Hungary	Rome	Milan	Naples	Madrid	Santiago
Germany		0.0669	<b>0.3440</b>	0.4468	0.6031	0.2112	0.9591	0.8063	0.3973	0.3319	0.1469	<b>0.0162</b>	0.3912
Barcelone			<b>0.0115</b>	0.1699	0.2988	0.4085	0.1366	0.1296	0.9436	0.3706	0.5696	<b>0.1030</b>	0.3079
Czech R.				<b>0.0454</b>	<b>0.0346</b>	<b>0.0013</b>	<b>0.0050</b>	<b>0.0182</b>	<b>0.2791</b>	<b>0.0938</b>	<b>0.0113</b>	<b>0.0016</b>	<b>0.3474</b>
Corunna					0.5984	0.5493	0.4308	0.2170	0.6016	0.6826	0.3159	<b>0.3074</b>	0.4058
Slovakia						0.3890	0.7542	0.3130	0.4279	0.7809	0.3860	<b>0.0240</b>	0.0481
Greece							0.0582	0.4206	0.3392	0.8381	0.2474	<b>0.2476</b>	0.3828
Netherlands								0.5251	0.2034	0.2601	0.4687	<b>0.0192</b>	0.1085
Hungary									0.7625	0.6050	0.0370	<b>0.0036</b>	0.0762
Rome										0.5897	0.2031	<b>0.0466</b>	0.4755
Milan											0.8025	<b>0.3435</b>	0.2648
Naples												<b>0.0143</b>	0.1335
Madrid													<b>0.1175</b>
Santiago													

Comparison was done with the exact test of population differentiation as implemented in Arlequin; global p value was  $< 10E-5$

The two collections from Greece were considered together as they were from the same town and did not differ in any of the studied SNPs.

Highlighted are the pair wise results of the Czech Republic and Madrid that concentrated most of the significant differences

Supplementary table 3

Pair-wise LD parameters in controls across the IRF5 locus

	rs729302	rs2004640	rs752637	rs13242262	rs10488630	rs10488631	rs2280714	rs4731535
rs729302		0.401	0.112	0.038	0.047	0.049	0.065	0.005
rs2004640	0.837		0.548	0.358	0.241	0.087	0.461	0.172
rs752637	0.363	0.909		0.607	0.199	0.058	0.759	0.37
rs13242262	0.218	0.707	0.809		0.368	0.064	0.729	0.62
rs10488630	0.394	0.676	0.75	0.985		0.073	0.281	0.534
rs10488631	0.976	0.987	0.982	1	1		0.048	0.122
rs2280714	0.26	0.919	0.959	0.975	0.981	1		0.469
rs4731535	0.095	0.438	0.784	0.979	0.954	0.989	0.969	

$r^2$

$D'$

Data from 13 sample collections (the Czech R. samples were excluded)



**Supplementary table 4**

Genotype frequencies at 8 IRF5 tag SNPs in each of the studied European sample collections

SNP	Group	Corunna			Barcelona			Greece			Germany			Netherlands			Hungary			Madrid			Milan			Naples			Rome			Santiago			Slovakia			
		0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2				
rs729302	control	%	48.4%	40.9%	10.8%	42.7%	46.9%	10.4%	48.1%	40.6%	11.3%	44.3%	42.9%	12.9%	47.5%	43.5%	9.0%	50.5%	42.1%	7.4%	47.5%	44.3%	8.2%	52.4%	41.0%	6.7%	42.5%	49.1%	8.5%	46.1%	41.6%	12.4%	44.4%	48.9%	6.7%	45.7%	39.4%	14.9%
	n	45	38	10	41	45	10	77	65	18	31	30	9	84	77	16	48	40	7	134	125	23	55	43	7	45	52	9	41	37	11	40	44	6	43	37	14	
	SLE	%	56.8%	40.9%	2.3%	52.2%	40.0%	7.8%	51.5%	41.8%	6.7%	49.4%	39.5%	11.1%	60.2%	37.9%	1.9%	48.9%	40.2%	10.9%	64.1%	32.6%	3.3%	40.6%	49.2%	10.2%	56.0%	39.6%	4.4%	54.8%	41.1%	4.1%	70.2%	26.9%	2.9%	63.8%	35.1%	1.1%
	n	50	36	2	47	36	7	69	56	9	40	32	9	62	39	2	45	37	10	59	30	3	52	63	13	51	36	4	40	30	3	73	28	3	60	33	1	
rs2004640	control	%	32.3%	48.4%	19.4%	34.4%	50.0%	15.6%	33.5%	46.0%	20.5%	24.3%	47.1%	28.6%	38.2%	43.9%	17.9%	31.9%	48.9%	19.1%	30.6%	50.0%	19.4%	35.2%	43.8%	21.0%	34.7%	47.4%	17.9%	31.8%	46.6%	21.6%	28.3%	55.4%	16.3%	26.6%	48.9%	24.5%
	n	30	45	18	33	48	15	54	74	33	17	33	20	66	76	31	30	46	18	85	139	54	37	46	22	33	45	17	28	41	19	26	51	15	25	46	23	
	SLE	%	40.9%	52.3%	6.8%	37.5%	50.0%	12.5%	40.3%	43.3%	16.4%	42.7%	35.4%	22.0%	41.0%	51.0%	8.0%	33.3%	46.2%	20.4%	45.7%	39.1%	15.2%	33.6%	48.4%	18.0%	42.7%	48.3%	9.0%	40.8%	49.3%	9.9%	53.8%	42.3%	3.8%	34.4%	50.5%	15.1%
	n	36	46	6	33	44	11	54	58	22	35	29	18	41	51	8	31	43	19	42	36	14	43	62	23	38	43	8	29	35	7	56	44	4	32	47	14	
rs752637	control	%	40.9%	46.2%	12.9%	54.2%	39.6%	6.3%	47.5%	41.8%	10.8%	34.3%	48.6%	17.1%	46.3%	43.4%	10.3%	49.5%	35.8%	14.7%	40.8%	49.6%	9.6%	43.8%	40.0%	16.2%	45.3%	42.5%	12.3%	49.4%	36.0%	14.6%	39.8%	41.0%	19.3%	39.4%	44.7%	16.0%
	n	38	43	12	52	38	6	75	66	17	24	34	12	81	76	18	47	34	14	115	140	27	46	42	17	48	45	13	44	32	13	33	34	16	37	42	15	
	SLE	%	48.9%	48.9%	2.3%	48.9%	41.1%	10.0%	49.2%	39.4%	11.4%	52.4%	40.2%	7.3%	47.6%	44.7%	7.8%	48.4%	37.6%	14.0%	52.2%	38.0%	9.8%	46.9%	44.5%	8.6%	48.3%	43.8%	7.9%	52.1%	42.5%	5.5%	59.6%	37.5%	2.9%	42.6%	45.7%	11.7%
	n	43	43	2	44	37	9	65	52	15	43	33	6	49	46	8	45	35	13	48	35	9	60	57	11	43	39	7	38	31	4	62	39	3	40	43	11	
rs13242262	control	%	43.0%	41.9%	15.1%	54.2%	35.4%	10.4%	47.7%	37.4%	14.8%	31.4%	51.4%	17.1%	45.8%	45.2%	9.0%	46.3%	35.8%	17.9%	39.0%	48.2%	12.8%	40.0%	41.9%	18.1%	35.2%	55.2%	9.5%	47.2%	37.1%	15.7%	42.2%	45.6%	12.2%	35.1%	48.9%	16.0%
	n	40	39	14	52	34	10	74	58	23	22	36	12	81	80	16	44	34	17	110	136	36	42	44	19	37	58	10	42	33	14	38	41	11	33	46	15	
	SLE	%	46.6%	47.7%	5.7%	41.1%	43.3%	15.6%	49.6%	40.9%	9.6%	53.1%	37.0%	9.9%	41.7%	48.5%	9.7%	48.4%	36.6%	15.1%	48.9%	42.4%	8.7%	45.2%	43.7%	11.1%	41.6%	49.4%	9.0%	56.2%	34.2%	9.6%	63.1%	34.0%	2.9%	39.4%	51.1%	9.6%
	n	41	42	5	37	39	14	57	47	11	43	30	8	43	50	10	45	34	14	45	39	8	57	55	14	37	44	8	41	25	7	65	35	3	37	48	9	
rs10488630	control	%	37.6%	45.2%	17.2%	28.1%	47.9%	24.0%	37.9%	46.6%	15.5%	45.7%	45.7%	8.6%	30.7%	53.4%	15.9%	42.1%	43.2%	14.7%	37.0%	45.6%	17.4%	35.2%	48.6%	16.2%	26.4%	63.2%	10.4%	34.4%	47.8%	17.8%	33.3%	51.6%	15.1%	41.5%	40.4%	18.1%
	n	35	42	16	27	48	23	61	75	25	32	32	6	54	94	28	40	41	14	104	128	49	37	51	17	28	67	11	31	43	16	31	48	14	39	38	17	
	SLE	%	50.0%	36.4%	13.6%	43.3%	46.7%	10.0%	48.4%	41.3%	10.3%	34.1%	43.9%	22.0%	32.4%	51.0%	16.7%	43.0%	46.2%	10.8%	31.5%	46.7%	21.7%	36.7%	48.4%	14.8%	43.8%	42.7%	13.5%	32.9%	50.7%	16.4%	37.5%	45.2%	17.3%	46.8%	42.6%	10.6%
	n	44	32	12	39	42	9	61	52	13	28	36	18	33	52	17	40	43	10	29	43	20	47	62	19	39	38	12	24	37	12	39	47	18	44	40	10	
rs10488631	control	%	78.5%	21.5%	0.0%	85.4%	13.5%	1.0%	80.1%	19.2%	0.6%	70.0%	28.6%	1.4%	71.0%	27.8%	1.1%	72.6%	25.3%	2.1%	85.5%	14.2%	0.4%	85.7%	13.3%	1.0%	85.8%	14.2%	0.0%	80.0%	18.9%	1.1%	78.9%	20.0%	1.1%	85.1%	14.9%	0.0%
	n	73	20	0	82	13	1	125	30	1	49	20	1	125	49	2	69	24	2	241	40	1	90	14	1	91	15	0	72	17	1	71	18	1	80	14	0	
	SLE	%	59.1%	35.2%	5.7%	68.9%	30.0%	1.1%	61.1%	33.6%	5.3%	79.3%	18.3%	2.4%	65.7%	30.4%	3.9%	71.0%	26.9%	2.2%	71.7%	26.1%	2.2%	63.0%	34.6%	2.4%	68.9%	30.0%	1.1%	64.4%	32.9%	2.7%	54.4%	40.8%	4.9%	60.6%	33.0%	6.4%
	n	52	31	5	62	27	1	69	38	6	65	15	2	67	31	4	66	25	2	66	24	2	80	44	3	62	27	1	47	24	2	56	42	5	57	31	6	
rs2280714	control	%	47.3%	41.9%	10.8%	60.0%	35.8%	4.2%	48.7%	39.9%	11.4%	37.1%	50.0%	12.9%	53.8%	40.4%	5.8%	54.3%	31.9%	13.8%	45.2%	45.9%	8.9%	49.5%	38.1%	12.4%	46.2%	47.2%	6.6%	55.6%	35.6%	8.9%	52.2%	41.1%	6.7%	43.6%	44.7%	11.7%
	n	44	39	10	57	34	4	77	63	18	26	35	9	92	69	10	51	30	13	127	129	25	52	40	13	49	50	7	50	32	8	47	37	6	41	42	11	
	SLE	%	51.1%	46.6%	2.3%	53.6%	39.3%	7.1%	49.6%	40.6%	9.8%	57.3%	35.4%	7.3%	52.0%	38.8%	9.2%	53.8%	34.4%	11.8%	54.3%	38.0%	7.6%	48.8%	43.3%	7.9%	49.5%	46.2%	4.4%	57.5%	37.0%	5.5%	64.7%	32.4%	2.9%	46.8%	43.6%	9.6%
	n	45	41	2	45	33	6	66	54	13	47	29	6	51	38	9	50	32	11	50	35	7	62	55	10	45	42	4	42	27	4	66	33	3	44	41	9	
rs4731535	control	%	33.3%	39.8%	26.9%	41.7%	42.7%	15.6%	29.7%	49.4%	20.9%	12.9%	54.3%	32.9%	24.9%	56.5%	18.6%	28.4%	40.0%	31.6%	30.5%	49.6%	19.9%	27.6%	52.4%	20.0%	27.6%	56.2%	16.2%	33.3%	43.3%	23.3%	29.7%	47.3%	23.1%	25.5%	51.1%	23.4%
	n	31	37	25	40	41	15	47	78	33	9	38	23	44	100	33	27	38	30	86	140	56	29	55	21	29	59	17	30	39	21	27	43	21	24	48	22	
	SLE	%	23.9%	45.5%	30.7%	18.9%	54.4%	26.7%	29.0%	41.9%	29.0%	39.0%	42.7%	18.3%	19.6%	52.9%	27.5%	29.0%	46.2%	24.7%	31.5%	45.7%	22.8%	32.8%	44.5%	22.7%	26.7%	46.7%	26.7%	31.5%	45.2%	23.3%	29.8%	45.2%	25.0%	17.0%	47.9%	35.1%
	n	21	40	27	17	49	24	36	52	36	32	35	15	20	54	28	27	43	23	29	42	21	42	57	29	24	42	24	23	33	17	31	47	26	16	45	33	

**Estudio de la regulación de la  
expresión génica de *IRF5* mediante  
datos de expresión procedentes de  
estudios con *microarrays*.**

RESEARCH ARTICLE

Open Access

# *Cis*-regulation of *IRF5* expression is unable to fully account for systemic lupus erythematosus association: analysis of multiple experiments with lymphoblastoid cell lines

Elisa Alonso-Perez<sup>1†</sup>, Marian Suarez-Gestal<sup>1†</sup>, Manuel Calaza<sup>1</sup>, Tony Kwan<sup>2</sup>, Jacek Majewski<sup>2</sup>, Juan J Gomez-Reino<sup>1,3</sup> and Antonio Gonzalez<sup>1\*</sup>

## Abstract

**Introduction:** Interferon regulatory factor 5 gene (*IRF5*) polymorphisms are strongly associated with several diseases, including systemic lupus erythematosus (SLE). The association includes risk and protective components. They could be due to combinations of functional polymorphisms and related to *cis*-regulation of *IRF5* expression, but their mechanisms are still uncertain. We hypothesised that thorough testing of the relationships between *IRF5* polymorphisms, expression data from multiple experiments and SLE-associated haplotypes might provide useful new information.

**Methods:** Expression data from four published microarray hybridisation experiments with lymphoblastoid cell lines (57 to 181 cell lines) were retrieved. Genotypes of 109 *IRF5* polymorphisms, including four known functional polymorphisms, were considered. The best linear regression models accounting for the *IRF5* expression data were selected by using a forward entry procedure. SLE-associated *IRF5* haplotypes were correlated with the expression data and with the best *cis*-regulatory models.

**Results:** A large fraction of variability in *IRF5* expression was accounted for by linear regression models with *IRF5* polymorphisms, but at a different level in each expression data set. Also, the best models from each expression data set were different, although there was overlap between them. The SNP introducing an early polyadenylation signal, rs10954213, was included in the best models for two of the expression data sets and in good models for the other two data sets. The SLE risk haplotype was associated with high *IRF5* expression in the four expression data sets. However, there was also a trend towards high *IRF5* expression with some protective and neutral haplotypes, and the protective haplotypes were not associated with *IRF5* expression. As a consequence, correlation between the *cis*-regulatory best models and SLE-associated haplotypes, regarding either the risk or protective component, was poor.

**Conclusions:** Our analysis indicates that although the SLE risk haplotype of *IRF5* is associated with high expression of the gene, *cis*-regulation of *IRF5* expression is not enough to fully account for *IRF5* association with SLE susceptibility, which indicates the need to identify additional functional changes in this gene.

**Keywords:** systemic lupus erythematosus *IRF5*, lymphoblastoid cell lines, *cis*-regulation, disease susceptibility, linear regression models

\* Correspondence: antonio.gonzalez.martinez-pedrayo@sergas.es

† Contributed equally

<sup>1</sup>Laboratorio Investigacion 10 and Rheumatology Unit, Instituto de Investigacion Sanitaria-Hospital Clinico Universitario de Santiago, Travesia Choupana sn, Santiago de Compostela E-15706, Spain

Full list of author information is available at the end of the article

## Introduction

Systemic lupus erythematosus (SLE) [1-4], Sjögren's syndrome [5-7], systemic sclerosis [8-11] and primary biliary cirrhosis [12,13] are complex autoimmune diseases with a genetic component that includes among their strongest susceptibility loci the interferon regulatory factor 5 gene (*IRF5*). There are reports indicating that this gene can be associated with a subgroup of patients with rheumatoid arthritis [14-16] and patients with other autoimmune diseases [17-19]. The *IRF5* gene encodes a transcription factor involved in the innate immune response as part of the type I IFN pathway, and its risk alleles have been associated with increased expression of this pathway [20,21]. Multiple polymorphisms in *IRF5* are associated with disease susceptibility, but it is unclear which of them is causal and how these polymorphisms contribute to disease predisposition. This uncertainty is a serious obstacle to progress in these complex diseases.

Four polymorphisms with a putative functional role have been described. One of them is an insertion-deletion polymorphism (indel) changing 10 amino acids in exon 6, but experimental evidence of any effect associated with this indel is still lacking [22,23]. The other three polymorphisms are involved in processes that could influence expression levels of *IRF5*. The T allele of rs2004640 introduces a donor splice site that exchanges alternative first exons. It could affect levels of *IRF5* mRNA through differences in *cis*-regulation [2], but its relevance has been questioned [22]. The CGGGG indel modulates binding of the Sp1 transcription factor in the *IRF5* promoter [24], but it did not contribute independently to *IRF5* levels in a study involving blood cells from healthy controls [25]. The strongest evidence of a role in *cis*-regulation has been found for the remaining functional polymorphism, rs10954213. Its A allele creates an early polyadenylation site that leads to a shorter mRNA isoform with an extended half-life and higher *IRF5* expression in both lymphoblastoid cell lines (LCLs) [3,23] and blood cells [25]. However, according to studies done with LCLs, this SNP is not enough to fully account for *IRF5* *cis*-regulation [3,23]. In addition, researchers in a study analysing *IRF5* expression in blood cells from SLE patients did not find any significant effect of this SNP or of any of the functional polymorphisms [26]. These contrasting pieces of evidence do not allow for a clear understanding of *IRF5* *cis*-regulation and its relationship to disease susceptibility.

*IRF5*-dependent disease susceptibility is determined by haplotypes with opposed effects: risk and protection [3-5,11,15,16,22,23]. The risk haplotype, identified by the rare allele of rs10488631 (or rs2070197), could be due to a combination of effects of the known functional

polymorphisms, but its components are unclear. It has been proposed to result from the combination of two functional polymorphisms, rs2004640 and rs10954213, and a SNP of unknown relevance [3], or from a gradation of the effects of three functional polymorphisms, the two mentioned plus the exon 6 indel [23], or from an epistatic interaction between a unique combination of alleles at the same three functional polymorphisms [4]. Other studies have left this matter more or less undefined because of the lack of convincing evidence of the relevance of all the polymorphisms' segregating with the risk haplotype [22], or they have proposed, after the discovery of the CGGGG indel, that this functional polymorphism determines SLE increased risk together with not yet known functional polymorphisms [24]. The protective haplotypes are represented by the rare allele of rs729302 that is 5' to the gene, but are not correlated with any of the functional polymorphisms [3,4,23]. Therefore, none of the two effects has a clear relationship to known functional polymorphisms or to *IRF5* function.

Here we address these questions using, for the first time, information from multiple mRNA expression studies and from the four known functional *IRF5* polymorphisms. This approach allowed us to assess the reproducibility and generality of the results. Also, it afforded us the opportunity to test the independent contribution of each functional polymorphism and to identify the SNP introducing an early polyadenylation signal, rs10954213, as the clearest *cis*-regulatory one. In addition, we have confirmed that the SLE risk haplotype is associated with high *IRF5* expression. However, the lack of correlations between *cis*-regulatory polymorphisms and SLE association and between *IRF5* expression and SLE protective haplotypes indicates that SLE association involves changes in *IRF5* function apart from its expression.

## Materials and methods

### Lymphoblastoid cell line expression data

*IRF5* expression data from two collections of LCLs were obtained from four published microarray studies (Table 1). Three of the studies were done with LCLs from unrelated subjects derived from the European population from the International HapMap Project (CEU) [27-29], which lacks significant admixture and has been used as the reference for the European Caucasian population in many studies. The fourth study was done with LCLs from children with asthma [30]. That study included 206 UK families with negligible population stratification. We have used only the LCLs from each family having the best genotype call rate, leaving us with a total of 181. Data were obtained from the Gene Expression Omnibus repository [27,28] (accession numbers GSE6536 and GSE2552) or from the study

**Table 1 Expression profiling studies in lymphoblastoid cell lines whose *IRF5* data have been analysed<sup>a</sup>**

Study	Number of LCLs	Microarray system used	Number of <i>IRF5</i> probes used	LCL collection group
Kwan <i>et al.</i> [29]	57	GeneChip Human Exon 1.0 ST Array (Affymetrix, Inc.)	17	CEU
Stranger <i>et al.</i> [28]	60	Illumina WG-6v1 BeadChip Array (Illumina, Inc.)	2	CEU
Cheung <i>et al.</i> [27]	58	GeneChip Human Genome Focus Array (Affymetrix, Inc.)	1	CEU
Dixon <i>et al.</i> [30]	181 <sup>b</sup>	GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix, Inc.)	3	Asthma

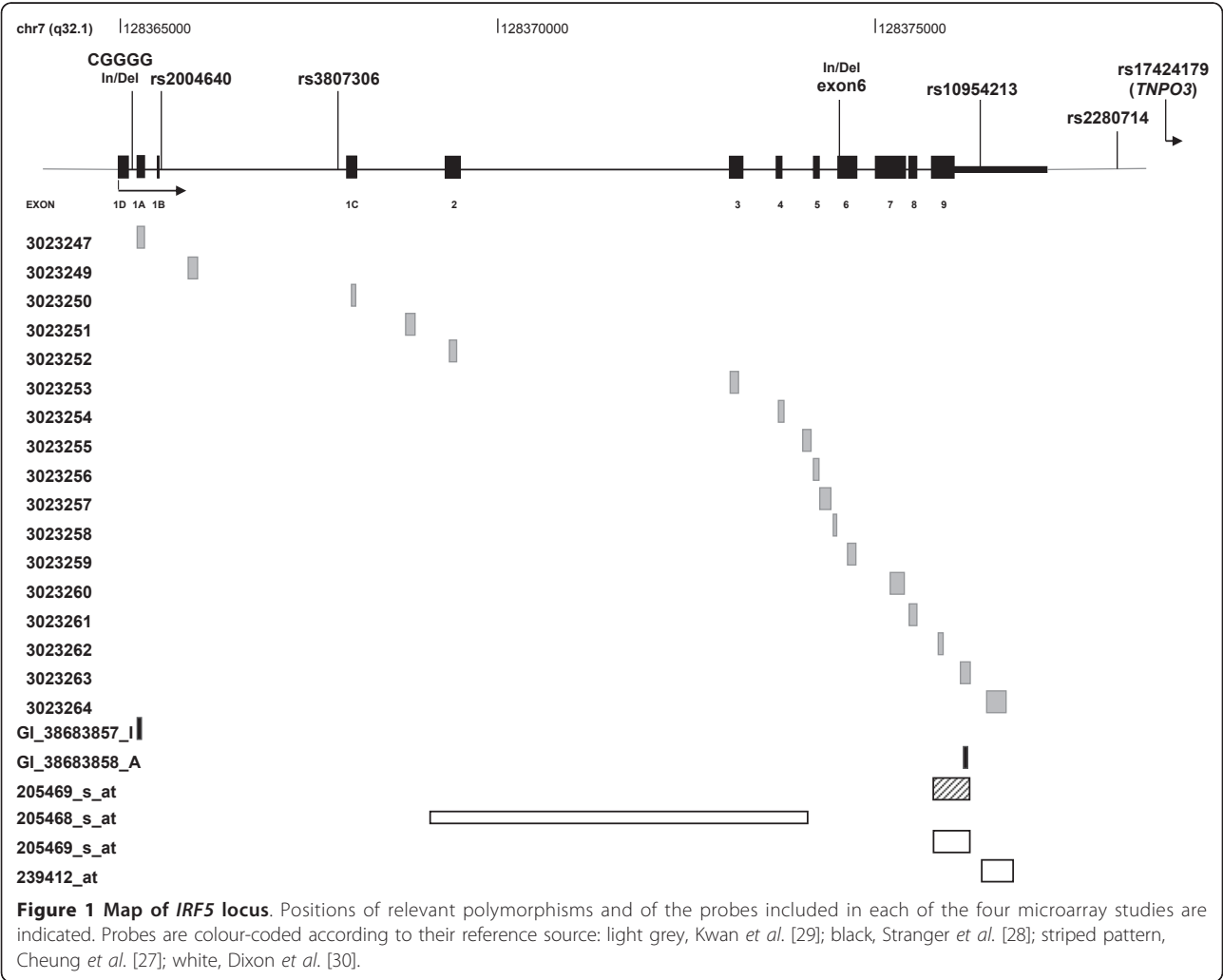
<sup>a</sup>*IRF5* = interferon regulatory factor 5 gene; LCL = lymphoblastoid cell line; CEU = European population from International HapMap Project database; <sup>b</sup>number of LCLs selected for having the best genotyping call rate per family among the 400 LCLs available.

authors [29,30]. Each study used a different microarray that included a variety of probes to examine *IRF5* expression (Figure 1). We used expression data only from validated probes in each study.

***IRF5* genotypes of the lymphoblastoid cell line**

The linkage disequilibrium (LD) block encompassing *IRF5* was defined according to International HapMap

Project data on the CEU population between 128,158 kb and 128,304 kb on chromosome 7 (HapMap Rel 21a/phase II Jan07, NCBI B35, dbSNPb125). Genotypes of the CEU LCLs for the 72 SNPs included in this 146-kb region (Additional file 1, Table S1) were downloaded from the International HapMap Project (HapMap) [31]. Data from 27 SNPs in this LD block were available in the asthma collection of LCLs.



### Genotyping and imputation of additional genotypes

We obtained complete genotype information in the *IRF5* LD block to a total of 109 polymorphisms (Additional file 1, Table S1) by imputation using MACH 1.0 software [32]. Information for imputation was taken from three sources: the 72 SNPs that had been studied in HapMap LCLs, the 27 SNPs available in the asthma LCLs and the 56 SNPs that we have genotyped in 95 healthy Spanish donors. In this way, it was possible to include the four putative functional polymorphisms (absent from HapMap) and more SNPs in the 5' region of *IRF5* (seven SNPs in tight LD with rs729302, the tagging SNP for the protector haplotypes in SLE) (Additional file 2, Figures S1 and S2). There were overlaps between the HapMap and asthma data sets (21 SNPs) and between HapMap and the Spanish donors (nine tagging SNPs). Genotypes of the Spanish donors were obtained by using the ABI PRISM SNaPshot Multiplex Kit (Applied Biosystems, Carlsbad, CA, USA) as described previously [4], except for rs3778752, rs3778751 and the CGGGG indel, which were sequenced (Additional file 1, Tables S2 and S3), and the exon 6 indel that was genotyped by length variation in agarose electrophoresis as described previously [4]. Polymorphisms with a MACH 1.0 quality score < 0.8 were discarded.

DNA samples from controls were obtained with their informed written consent, and the study was approved by the Committee for Clinical Research of Galicia (Spain).

### Statistical analysis

Expression results from probes targeting introns were excluded from the analysis. Expression data were transformed into standardised normal distributions (that is, expression data from each probe were transformed into new variables with mean = 0 and standard deviation = 1 by subtracting the mean to each value and dividing the result by the standard deviation) to avoid differences in scale when performing comparisons between studies. Relations of the expression results with the *IRF5* polymorphisms were analysed by multiple linear regression. These analyses were performed with a forward entry procedure, which adds new polymorphisms to the regression model one-by-one, starting with the most associated until no further significant improvement is achieved or until one of the polymorphisms does not show a significant contribution to the model. A genetic additive model (with values 0, 1 and 2 for the AA, Aa and aa genotypes, respectively) was considered. Only one of each pair of polymorphisms, to a total of 35 polymorphisms, with  $r^2 \geq 0.90$  was included in the analyses to avoid collinearity problems (Additional file 1, Table S1). Nested linear regression models were compared using the likelihood ratio test. Nonnested models were compared using

Davidson and MacKinnon's J-test [33], which specifies a proxy parameter in an artificial nested model combining the two nonnested models and then tests the proxy parameter. All analyses were done using Statistica 7.0 software (StatSoft, Tulsa, OK, USA) or in R software implementations, except for haplotype estimation, which was done using Phase 2 software [34].

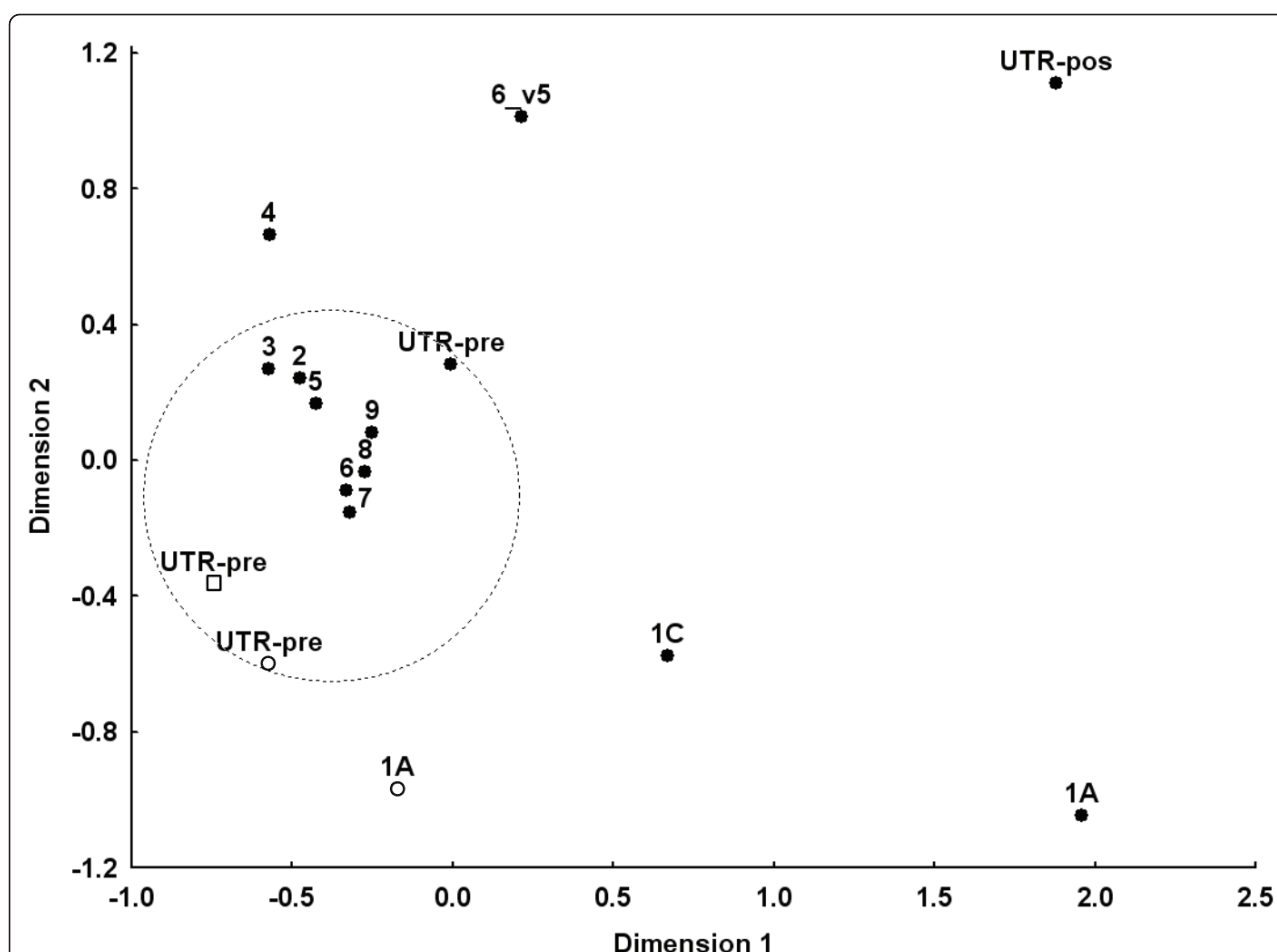
### Results

#### *IRF5* expression in lymphoblastoid cell lines

To ascertain *cis*-regulatory *IRF5* polymorphisms, we selected four studies (Table 1) containing *IRF5* genotypes and microarray expression data in LCLs [27-30]. The multiplicity of studies and hybridisation probes (Figure 1) allowed us to select the most representative expression results. As a first step in this process, we used the study by Kwan *et al.* [29], which included 13 probes targeting specific *IRF5* exons in LCLs from the CEU population of HapMap. Two different groups of results were identified (Figure 2). The first group included eight probes that were highly correlated (mean pairwise  $r^2 = 0.79$ ). They hybridised with exons 2, 3, 5, 6 (not including the 30-bp indel), 7, 8 and 9 and with the 3'UTR previous to the early polyadenylation signal SNP rs10954213. The uncorrelated results were obtained with probes hybridising with exons 1A and 1C, which are alternatively spliced and untranslated; with exon 4, which is the smallest; and with the sequence of exon 6, which is present only in splice variant 5. We took the average of the first group as representative of *IRF5* expression and named it K8.

The other two microarray studies done with CEU LCLs contained fewer *IRF5* probes. The Stranger *et al.* study [28] included probes hybridising with exon 1A and with the 3'UTR previous to rs10954213 (Figure 1). Only results from the second probe correlated with K8 ( $r^2 = 0.56$ ), and they were taken as representative and called S (Figure 2). The Cheung *et al.* study [27] included only one *IRF5* probe (Figure 1), which hybridised with the upstream region of exon 9 and the 3'UTR. The results of this probe, which we identified as C, strongly correlated with the results for K8 ( $r^2 = 0.60$ ) and S ( $r^2 = 0.75$ ) (Figure 2). The high correlation between the three data sets, K8, S and C, permitted us to obtain a global average that was denoted KSC. For an analysis of the unselected data, see Additional file 2, Supplementary Information.

To increase the generality of the results, we used a fourth study that had examined a different collection of LCLs [30]. That study included data derived from three *IRF5* probes (Figure 1), which were poorly correlated (not shown). We considered as representative only the probe that was shared with the study by Cheung *et al.* [27] and targeted sequences addressed in the other two studies. These data were named D.



**Figure 2 Correlation between *IRF5* expression results obtained with different probes and in different experiments.** Only results obtained with the same European population from the International HapMap Project (CEU) lymphoblastoid cell line (LCL) are compared. Bidimensional, nonparametric, multidimensional scaling was used for representation. Data from probes selected as representative are within the dashed circle. Labels for data from each probe indicate the number of the exon they target, including the 1A and 1C alternative exons and a variant sequence of exon 6 (6\_v5), and UTR-pre and UTR-pro for the probes targeting the 3'UTR previous or posterior to rs10954213, respectively. Filled circles correspond to data from Kwan *et al.* [29], empty circles correspond to data from Stranger *et al.* [28] and empty squares correspond to data from Cheung *et al.* [27].

### Best genetic models of *IRF5* cis-regulation

The forward entry multiple linear regression process led to the identification of four best models, one for each of the data sets and a best model for the KSC average that included the same polymorphisms as the best model for S. The best model for the K8 results, which included genotypes at two SNPs, rs3807306 and rs17424179 (Table 2), explained 0.31 of the variance in *IRF5* expression. The first SNP, rs3807306, is in *IRF5* intron 1 and is the most associated in the model. This SNP showed an  $r^2$  value larger than 0.7 with 22 polymorphisms, including two functional ones: rs10954213 ( $r^2 = 0.79$ ) and the CGGGG indel ( $r^2 = 0.75$ ). The second SNP, rs17424179, is 68 kb 3' to *IRF5*. It did not show a strong correlation with any other polymorphism (all pairwise  $r^2 < 0.2$ ). Its contribution to the model fit was small. The best model for S results

included three SNPs that accounted for a very large fraction of variability in *IRF5* expression (adjusted  $r^2 = 0.80$ ). The strongest association in this model was with rs10954213 (Table 2). The two other SNPs were the same as the best model for K8, rs3807306 and rs17424179. The best model for the C expression data also accounted for a large fraction of variability (adjusted  $r^2 = 0.55$ ), including only two SNPs (Table 2). The major contribution corresponded to rs2280714, which is 4.6 kb 3' to *IRF5* and showed a strong correlation with the second SNP in the model, rs10954213 ( $r^2 = 0.84$ ), and with 25 other SNPs ( $r^2 > 0.7$ ). As this model includes two highly correlated SNPs, their independent contributions were severely reduced in relation to the model fit ( $P = 0.02$  for each of the two SNPs in a model with  $P = 9.2 \times 10^{-11}$ ). As a form of summary of these three data sets, the average KSC



**Table 2 Best multiple linear regression models with *cis*-polymorphisms accounting for *IRF5* gene expression in each of the four data sets and in the average expression from CEU LCL (KSC)<sup>a</sup>**

	Data set	Best linear regression model		Polymorphism <i>P</i> value in model	
		Adjusted $r^2$	Model <i>P</i>	Polymorphisms	
K8 (Kwan et al. [29])	0.31	$1.7 \times 10^{-5}$	rs3807306	$4.0 \times 10^{-6}$	
			rs17424179	0.026	
S (Stranger et al. [28])	0.80	$2.5 \times 10^{-20}$	rs10954213	$1.2 \times 10^{-4}$	
			rs3807306	$2.4 \times 10^{-3}$	
			rs17424179	$9.5 \times 10^{-3}$	
C (Cheung et al. [27])	0.55	$9.2 \times 10^{-11}$	rs2280714	0.02	
			rs10954213	0.02	
KSC	0.69	$1.3 \times 10^{-13}$	rs10954213	$5.1 \times 10^{-3}$	
			rs3807306	0.013	
			rs17424179	0.011	
D (Dixon et al. [30])	0.28	$1.3 \times 10^{-13}$	CGGGG indel	$2.4 \times 10^{-6}$	
			rs2280714	$8.3 \times 10^{-4}$	

<sup>a</sup>*IRF5* = interferon regulatory factor 5 gene; LCL = lymphoblastoid cell line; CEU = European population from International HapMap Project database; KSC = average of standard normal transformed K8, S and C data. Statistical parameters of the best models are provided together with *P* values corresponding to independent contribution of each polymorphism to the model. Data sets in left column are representative *IRF5* expression results.

expression results were analysed. They were well accounted for (66% of the variability) by a best model with three SNPs that were the same and in the same order as those in the best model for S data (Table 2). Therefore, the three studies with the same cell lines showed expression data that could be largely explained by *cis*-regulation because of a small number of polymorphisms.

The best model for the independent D results comprised two polymorphisms: the functional CGGGG indel and the already mentioned rs2280714 (Table 2), which is strongly correlated with rs10954213, among others. The polymorphism composition of the best genetic models was applied to the other expression data sets to explore their relationships. The exchanged models were significantly inferior to the proper best models, with two exceptions (Figure 3): the model with the best combination of polymorphisms for the S data set was equivalent to the best model in the K8 data, which it contained; and the model with the best combination of polymorphisms for the D data set was equivalent to the best model in the C expression data. The SNP composition rs10954213, rs3807306 and rs17424179 produced the best model overall: best for the S data set, not significantly different from the best in the K8 data set, second best for the D data and third best for the C data set. In addition, it comprised the SNPs in the best model for the KSC average data.

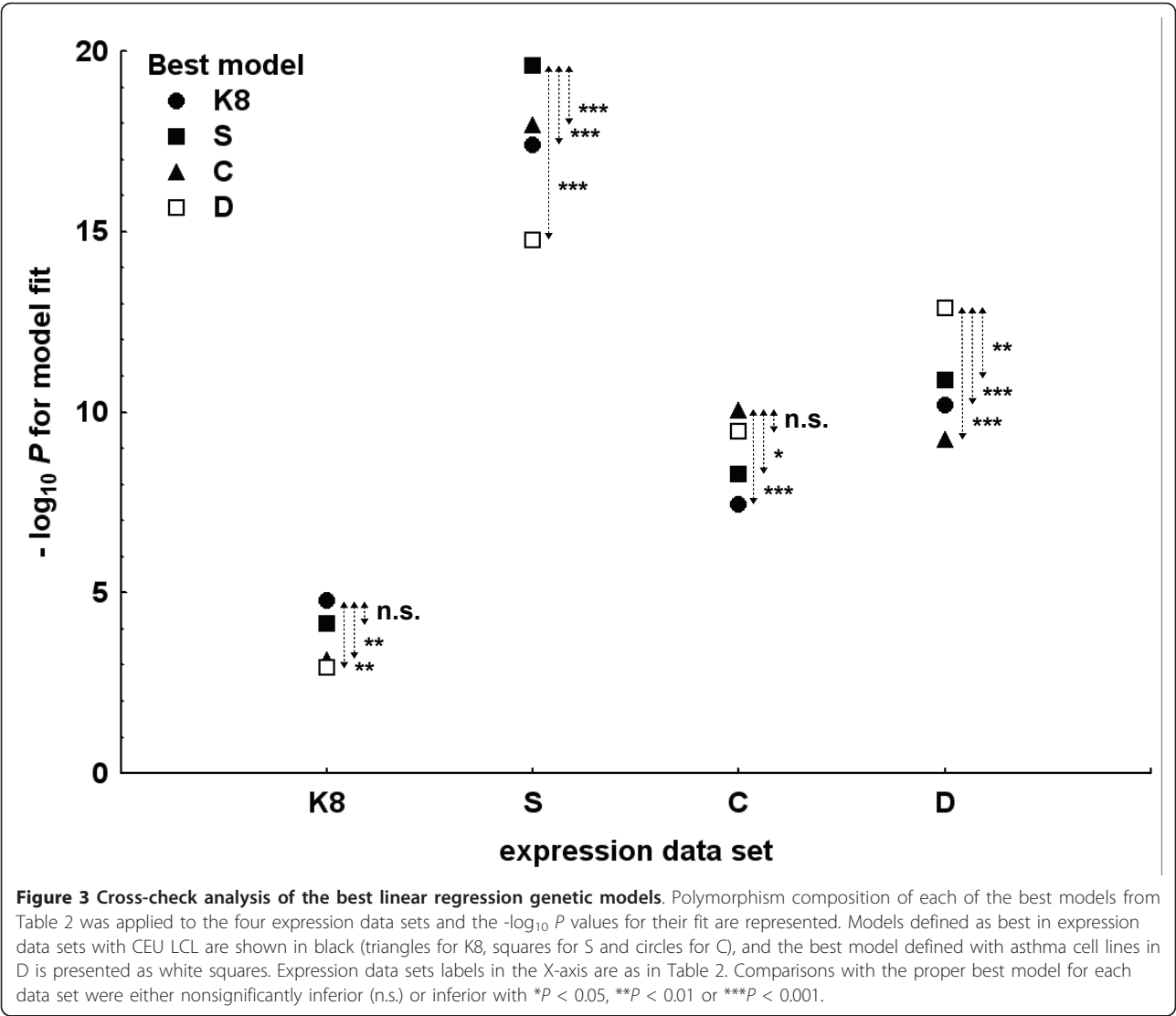
To ascertain the origin of differences between data sets, we compared model fit with the scale of expression levels, the range of values and their dispersion, as well as with sample size differences, but no correlation was found. Also, we assessed the best model for K8 in the results from each of the eight exons in the study by

Kwan et al. [35]. The fit of the model ranged from  $r^2 = 0.12$  for data from exon 9 and 3'UTR (previous to rs10954213) to  $r^2 = 0.37$  for data from exon 7 ( $P = 0.01$ , and  $1.7 \times 10^{-6}$  in linear regression analyses, respectively). These two exons are shared by all known *IRF5* isoforms. Therefore, these differences point to the probes as the source of variability because the results are from the same hybridisation experiment, cancelling all variation involved in cell culture, mRNA extraction and cDNA synthesis or labelling. In contrast, data sets C and D shared the same probe, but they also showed differences in model fit that should be ascribed, in this case, to other unidentified factors that could include laboratory procedures and the collection of cells from healthy subjects and asthma patients in C and D set, respectively.

#### Role of the putative functional polymorphisms

We have analysed how well models in which only functional polymorphisms were included accounted for the expression data (Figure 4). Models including the exon 6 indel were clearly inferior and are not shown, given the lack of any previous evidence of the involvement of the exon 6 indel in *IRF5* *cis*-regulation. Each of the remaining three functional polymorphisms, considered individually, was significantly associated with *IRF5* expression in the four data sets. The early polyadenylation signal SNP, rs10954213, was clearly dominant among the functional polymorphisms in these individual comparisons, except in the D data set. However, none of them alone was able to account for *IRF5* expression equivalently to the proper best genetic model for each data set. Models combining functional polymorphisms were not better than models with rs10954213 alone in the three data





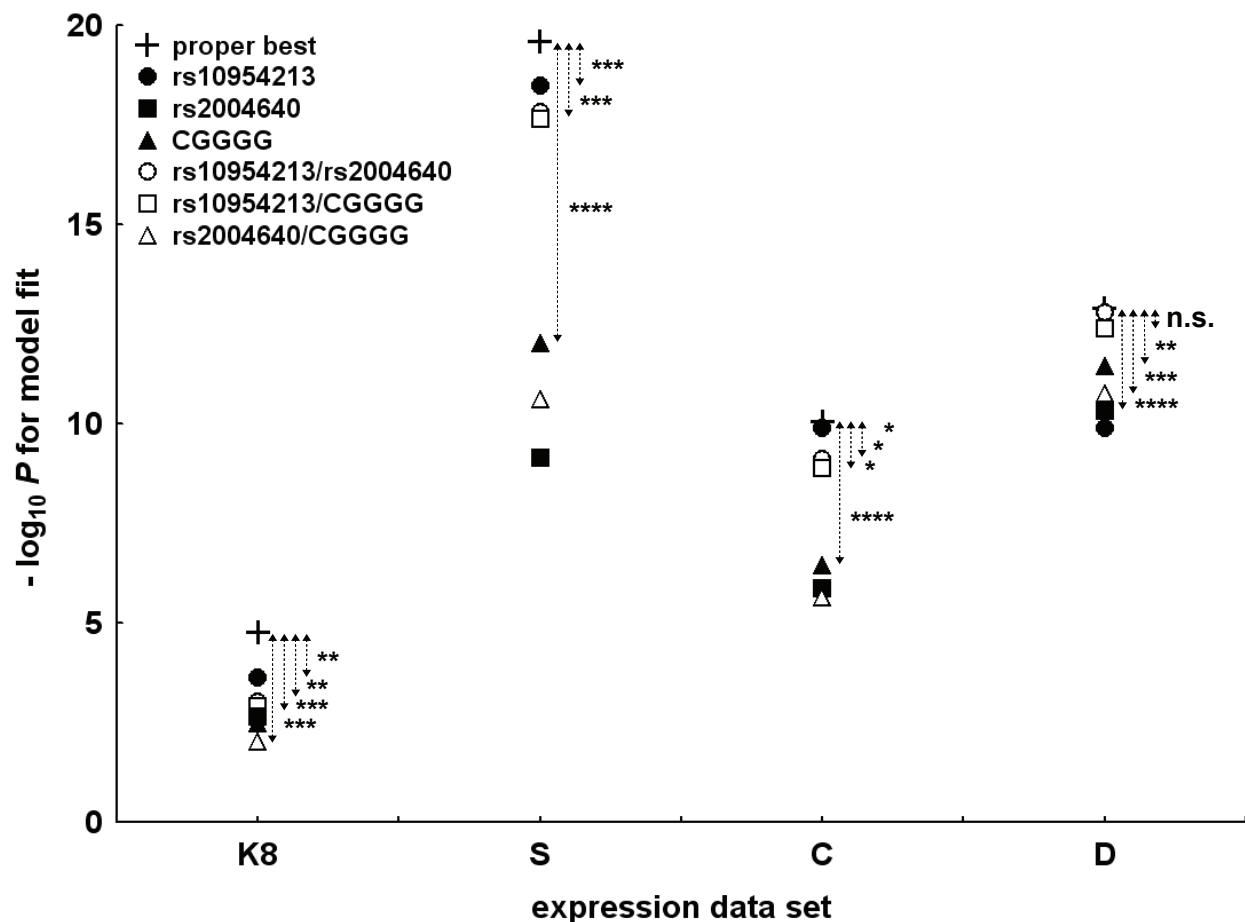
sets with CEU LCLs (K8, S and C). In contrast, the two models including rs10954213, together with rs2004640 or with the CGGGG indel, were the best in accounting for the D expression data and were not inferior to the proper best model. It is important to note that in these two models and in this data set, the two component polymorphisms showed a significant independent contribution (Additional file 2, Table S1). In the other three data sets (K8, S and C), the best models with two functional polymorphisms included rs10954213 with rs2004640 and, immediately below, rs10954213 with the CGGGG indel. Only rs10954213, however, showed a significant independent contribution in these combined models (Additional file 2, Table S1).

A search for other putative functional polymorphisms in the *IRF5* sequence with two bioinformatics applications, Pupasuite 3.1 [36] and FastSNP [37], gave only an

SNP that could introduce an alternative splice site in intron 1, but it was not polymorphic in our 95 Spanish samples.

#### Relationship between *IRF5* expression and systemic lupus erythematosus susceptibility

We used haplotypes defined in previous reports to assess the relationship between *IRF5* expression and SLE susceptibility [2,4] (Additional file 2, Table S2). The SLE risk haplotype H6 is identified by the minor allele of rs10488631. The protective haplotypes H1 and H2 are defined by the minor allele of rs729302, with H1 including the A allele of rs10954213 and H2 including the G allele. They share the minor allele of rs2004640 with the neutral haplotype H3, but this latter haplotype lacks the minor allele of rs729302. H4 and H5, which are SLE-neutral, are very similar to the risk haplotype H6 but



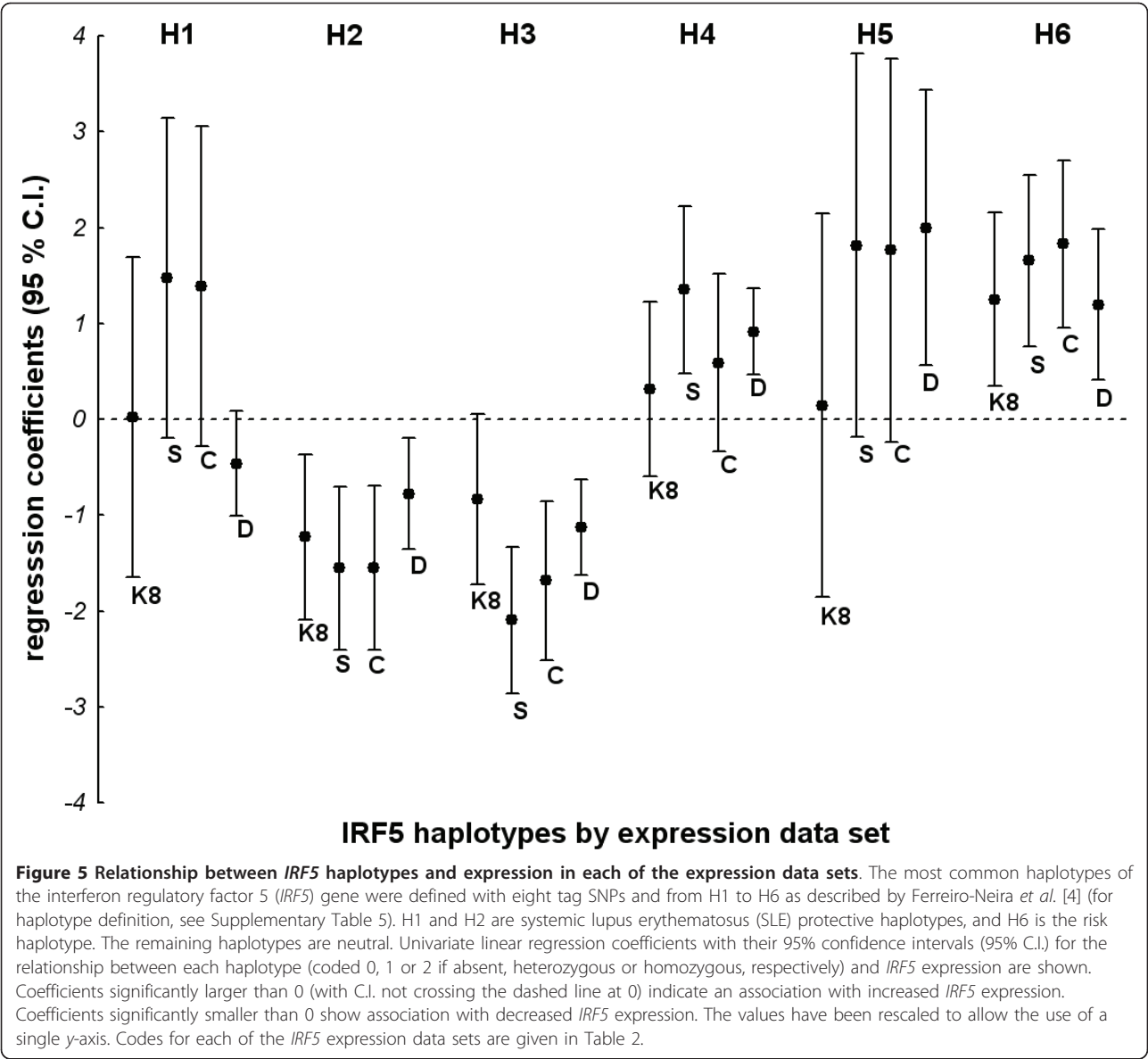
**Figure 4 Evaluation of linear regression models made only of known functional polymorphisms.** The  $-\log_{10} P$  value for the fit of the models applied to each data set are represented in ordinates and compared with the proper best models (plus signs). Models including only one of the functional polymorphisms are indicated by filled symbols, and models combining two polymorphisms are shown as open symbols. Models included either rs10954213 (filled circles), rs2004640 (filled squares), the CGGGG insertion-deletion polymorphism (indel) (filled triangles), rs10954213 and rs2004640 (open circles), rs10954213 and the CGGGG indel (open squares) or the rs2004640 and the CGGGG indel (open triangles). Comparisons with the proper best model for each data set were either nonsignificantly inferior (n.s.) or inferior with \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  or \*\*\*\* $P < 0.0001$ . Expression data sets in the X-axis are as in Table 2.

lack the minor allele of rs10488631. None of the best models for any of the expression data sets was strongly correlated (all  $r^2 \leq 0.15$ ) with the haplotypes defining SLE risk (H6) or SLE protection (H1 and H2) (Additional file 2, Table S3). However, we found that the minor allele of rs10488631 that identifies the SLE risk haplotype was significantly associated with increased *IRF5* expression in all data sets (all  $P < 0.0094$ ). On the contrary, the minor allele of rs729302 that identifies the SLE protection haplotype was associated with lower expression of *IRF5* only in the D data set ( $P = 0.002$ ), but not in the other data sets (not shown). In addition, analysis of the association of the estimated haplotypes showed that the only haplotype consistently associated with high *IRF5* expression in all data sets was the risk haplotype H6 (Figure 5). However, this finding was not

specific, because there was an association of higher *IRF5* expression with neutral haplotypes H4 and H5 in some data sets. There was also poor correlation between SLE-protective haplotypes and *IRF5* expression. Of the two protective haplotypes, only the H2 haplotype was consistently associated with lower *IRF5* expression (all  $P$  values within the range from 0.009 to 0.0008) (Figure 5). The H1 haplotype was not associated with *IRF5* expression in any of the four data sets (all  $P > 0.09$ ).

## Discussion

Identification of SLE causal polymorphisms in *IRF5* is very difficult. A thorough analysis with novel characteristics including the use of expression data from four different studies, the inclusion of genotypes of the four known functional polymorphisms, and the direct



assessment of the relationship between SLE-associated haplotypes and *IRF5* expression has provided new and interesting insights.

The multiplicity of probes and studies allowed us to select the most representative *IRF5* expression data. They included almost all the probes for *IRF5* transcribed sequences that are common to all isoforms. They showed good correlation ( $r^2 \geq 0.56$ ) between the three experiments using the same LCLs. However, differences between experiments led to consequences in the results, such as with regard to the fraction of expression variability accounted for by the best *cis*-regulatory models, which ranged from 0.28 to 0.80. There is no simple cause of these differences. They did not correlate with sample size, with the scale of the expression values or

with their dispersion. However, there were notable differences within the same experiment that were dependent on the hybridisation probes. Other undefined factors, which could include cell culture, sample processing and differences between cell collections, were also suggested by our analyses.

Differences between the experiments were also evident in the best *cis*-regulatory models. Each expression data set was best explained by a specific genetic model, but the polymorphisms included in them were partially coincident. The best assessment of the relationships between the four best genetic models was obtained by applying the models to the other data sets as a sort of cross-checking procedure (Figure 3). This analysis showed that it was impossible to identify a single best

model explaining *IRF5* expression, but it was possible to define a range of best *cis*-regulatory models that were useful for assessing hypotheses.

The multiplicity of best *cis*-regulatory models makes it uninteresting to comment on the implications of each. However, there was a specific combination of SNPs worth discussing because it was superior to the others in cross-comparisons. It contained rs10954213 (see next paragraph) together with rs3807306 and rs17424179 (Figure 3). There has been no previous specific mention of rs17424179, but rs3807306 has been found to account for most haplotype effects in SLE association among African Americans [38], and it has been highlighted among the *IRF5* polymorphisms particularly associated with multiple sclerosis [18] and rheumatoid arthritis [14]. In addition, a model with rs3807306 and rs10954213 was the second best to account for SLE association in a study in Caucasians [24]. Some of these associations have been interpreted as if this SNP acted as a proxy for the CGGGG indel because of LD between them and lack of an allele-specific effect of rs3807306 in an electrophoretic mobility shift assay [18,24]. However, this interpretation is not applicable to our results, because the indel was also included in the analyses. Therefore, it is likely that rs3807306 indicates additional functional polymorphisms or a combination thereof. Its role in *IRF5* *cis*-regulation is also supported by the analysis of Rullo *et al.* [21], in which this SNP showed the strongest association with *IRF5* levels among the 14 SNPs considered, in both the European and the Asian collections of HapMap LCLs.

The SNP determining early *IRF5* polyadenylation, rs10954213, was clearly dominant in accounting for the expression data obtained with CEU LCLs (Table 2). This predominance of rs10954213 in the CEU LCLs was confirmed in the analyses limited to the known functional polymorphisms (Figure 4). Overall, our analyses provide strong evidence supporting the role of rs10954213 in *cis*-regulating *IRF5* expression. They are compatible with its identification as the main *IRF5* *cis*-regulatory polymorphism in blood cells from healthy controls [25] and with previous studies in which its role and mechanism of action were elucidated [2,3]. However, Feng *et al.* [26] recently reported a lack of association, a result that could be due to insufficient power because only 14 to 26 subjects were considered in these specific comparisons.

The dominant role of rs10954213 in the CEU LCL data was such that all the genetic models with known functional polymorphisms were inferior to the model including only rs10954213. These results suggest that the other three known functional polymorphisms are redundant in *IRF5* expression. This conclusion should be tempered by the discordant results obtained with the

asthma LCL expression data. They showed a significant contribution to the best functional *cis*-regulatory models from either rs2004640 or the CGGGG indel in combination with rs10954213. However, we do not know which of the results with the two LCL collections is more representative of the population at large.

We found a consistent association of the SLE risk haplotype with higher *IRF5* expression in the four data sets (Figure 5). This association has already been demonstrated in SLE and healthy control blood cells [25,26]. It has been the focus of attention in previous reports and is the basis of the hypothesis that *IRF5* risk alleles act by potentiating the type I IFN pathway. This hypothesis has received recent experimental support in studies done with SLE sera [20] and with LCLs [21].

Our results also indicate that there is more than *IRF5* expression in *IRF5*-dependent disease association. This was shown by the lack of correlation of the SLE susceptibility haplotypes with the best *cis*-regulatory models and with *IRF5* expression in either the SLE risk or SLE protective haplotype. We do not yet have a good hypothesis of what the additional changes in *IRF5*, besides its expression, could be. Possibilities include alteration of interactions with other proteins, as has been suggested for the exon 6 indel [22,23], or changes in the isoforms by alterations in splicing, a mechanism demonstrated for rs2004640 [2], but with little relevance [22]. A search for other putative functional polymorphisms using bioinformatics tools did not lead us to new hypotheses. Therefore, the need to continue studying *IRF5* polymorphisms to understand their role in disease susceptibility is an imperative.

One of the limitations of our study is that only global *IRF5* expression data, as opposed to isoform-specific data, were obtained from LCLs in basal conditions, which could be different from the relevant *IRF5* isoform, cell type or activation status. However, it is important to note that no significant *cis*-regulation for *IRF5* isoforms has yet been reported, in spite of its many splice variants and their upregulation in SLE [22,26]. In addition, results with blood cells have been concordant with results with LCLs [25], and the *IRF5* risk haplotype has also been found to be associated with overexpression of *IRF5* in the blood cells, monocytes and myeloid dendritic cells of SLE patients [26]. An additional limitation which we acknowledge is the possibility that some of the best models could be different with the use of actual genotypes in place of imputed ones. Finally, our study included only LCLs from European Caucasians. This was done on purpose because there are differences in the structure of *IRF5* haplotypes and their SLE associations and differences in *IRF5* *cis*-regulation between Europeans, Asians and Africans [21,38,39].

## Conclusions

Our study has shown significant variability in results from different studies of *IRF5* *cis*-regulatory polymorphisms. However, this variability is compatible with the finding that *cis*-regulatory changes in *IRF5* expression are not sufficient to explain their association with SLE, although there is a consistent association of the SLE risk haplotype with high *IRF5* expression.

## Additional material

**Additional file 1: Supplementary materials and methods.** Interferon regulatory factor 5 (*IRF5*) gene polymorphisms that have been studied, with indications of the sources of their expression data as well as the primers and probes that were used to genotype them.

**Additional file 2: Supplementary results.** Complementary analyses of the *IRF5* lineal regression models and of the haplotype distribution, together with linkage disequilibrium maps and expression results pertaining to probes targeting less representative *IRF5* exons.

## Abbreviations

CEU: European population from the International HapMap Project; CI: confidence interval; IFN: interferon; *IRF5*: interferon regulatory factor 5 gene; LCL: lymphoblastoid cell line; OR: odds ratio; SLE: systemic lupus erythematosus; SNP: single-nucleotide polymorphism; UTR: untranslated region.

## Acknowledgements

We thank Liming Liang of the Harvard School of Public Health (Boston, MA, USA) and William Cookson of Imperial College London (London, UK) for providing us with expression data and complementary information from their microarray study. EAP is the recipient of an Instituto de Salud Carlos III predoctoral bursary. MSG is the recipient of a Formacion de Profesorado Universitario predoctoral bursary from the Spanish Ministry of Education. MC is the recipient of an "Isabel Barreto" bursary from the Government of Galicia. This project was supported by grants PI06/0620 and PI080744 from the Instituto de Salud Carlos III (Spain) with funds from European Regional Development Fund (European Union).

## Author details

<sup>1</sup>Laboratorio Investigacion 10 and Rheumatology Unit, Instituto de Investigacion Sanitaria-Hospital Clinico Universitario de Santiago, Travesia Choupana sn, Santiago de Compostela E-15706, Spain. <sup>2</sup>Department of Human Genetics, McGill University, 1205 Dr Penfield Avenue, Montreal H3A 1B1, Canada. <sup>3</sup>Department of Medicine, University of Santiago de Compostela, San Francisco sn, Santiago de Compostela, E-15782, Spain.

## Authors' contributions

EAP genotyped the samples and participated in the interpretation of the results and the writing of the manuscript. MSG participated in the design of the study, obtained genotype data and participated in the interpretation of the results and the writing of the manuscript. MC participated in the design of the study, in statistical analysis and in the interpretation of the results. TK and JM provided detailed microarray data and participated in the interpretation of the results and the writing of the manuscript. JJGR participated in the analysis and interpretation of the results. AG participated in the design of the study and the acquisition of data and supervised the genotyping, statistical analysis, interpretation of results and the writing of the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 28 January 2011 Revised: 8 April 2011  
Accepted: 31 May 2011 Published: 31 May 2011

## References

1. Sigurdsson S, Nordmark G, Göring HH, Lindroos K, Wiman AC, Sturfelt G, Jönsen A, Rantapää-Dahlqvist S, Möller B, Kere J, Koskenmies S, Widén E, Eloranta ML, Julkunen H, Kristjansdóttir H, Steinsson K, Alm G, Rönnblom L, Syvänen AC: **Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus.** *Am J Hum Genet* 2005, **76**:528-537.
2. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, Ortmann WA, Koeuth T, González Escribano MF, Argentine and Spanish Collaborative Groups, Pons-Estel B, Petri M, Daly M, Gregersen PK, Martín J, Altschuler D, Behrens TW, Alarcón-Riquelme ME: **A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus.** *Nat Genet* 2006, **38**:550-555.
3. Cunningham-Graham DS, Manku H, Wagner S, Reid J, Timms K, Gutin A, Lanchbury JS, Vyse TJ: **Association of *IRF5* in UK SLE families identifies a variant involved in polyadenylation.** *Hum Mol Genet* 2007, **16**:579-591.
4. Ferreira-Neira I, Calaza M, Alonso-Perez E, Marchini M, Scorza R, Sebastiani GD, Blanco FJ, Rego I, Pullmann R Jr, Pullmann R, Kallenberg CG, Bijl M, Skopouli FN, Mavromati M, Migliaresi S, Barizzone N, Ruzickova S, Dostal C, Schmidt RE, Witte T, Papasteriades C, Kappou-Rigatou I, Endreffy E, Kovacs A, Ordi-Ros J, Balada E, Carreira P, Gomez-Reino JJ, Gonzalez A: **Opposed independent effects and epistasis in the complex association of *IRF5* to SLE.** *Genes Immun* 2007, **8**:429-438.
5. Miceli-Richard C, Comets E, Loiseau P, Puechal X, Hachulla E, Mariette X: **Association of an *IRF5* gene functional polymorphism with Sjögren's syndrome.** *Arthritis Rheum* 2007, **56**:3989-3994.
6. Miceli-Richard C, Gesteremann N, Ittah M, Comets E, Loiseau P, Puechal X, Hachulla E, Gottenberg JE, Lebon P, Becquemont L, Mariette X: **The CCGGG insertion/deletion polymorphism of the *IRF5* promoter is a strong risk factor for primary Sjögren's syndrome.** *Arthritis Rheum* 2009, **60**:1991-1997.
7. Nordmark G, Kristjansdóttir G, Theander E, Eriksson P, Brun JG, Wang C, Padyukov L, Truedsson L, Alm G, Eloranta ML, Jonsson R, Rönnblom L, Syvänen AC: **Additive effects of the major risk alleles of *IRF5* and *STAT4* in primary Sjögren's syndrome.** *Genes Immun* 2009, **10**:68-76.
8. Dieudé P, Guedj M, Wipff J, Avouac J, Fajardy I, Diot E, Granel B, Sibilia J, Cabane J, Mouthon L, Cracowski JL, Carpentier PH, Hachulla E, Meyer O, Kahan A, Boileau C, Allanore Y: **Association between the *IRF5* rs200640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis 2009.** *Arthritis Rheum* 2009, **60**:225-233.
9. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Hikami K, Kawamoto M, Fujimoto M, Takehara K, Sato S, Hara M, Tsuchiya N: **Association of a functional polymorphism in the *IRF5* region with systemic sclerosis in a Japanese population.** *Arthritis Rheum* 2009, **60**:1845-1850.
10. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van 't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, González-Gay MA, González-Escribano MF, Spanish Scleroderma Group, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, et al: **Genome-wide association study of systemic sclerosis identifies *CD247* as a new susceptibility locus.** *Nat Genet* 2010, **42**:426-429.
11. Dieudé P, Dawidowicz K, Guedj M, Legrain Y, Wipff J, Hachulla E, Diot E, Sibilia J, Mouthon L, Cabane J, Amoura Z, Cracowski JL, Carpentier P, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **Phenotype-haplotype correlation of *IRF5* in systemic sclerosis: role of 2 haplotypes in disease severity.** *J Rheumatol* 2010, **37**:987-992.
12. Hirschfield GM, Liu X, Han Y, Gorlov IP, Lu Y, Xu C, Chen W, Juran BD, Coltescu C, Mason AL, Milkiewicz P, Myers RP, Odin JA, Luketic VA, Speicene D, Vincent C, Levy C, Gregersen PK, Zhang J, Heathcote EJ, Lazaridis KN, Amos CI, Siminovich KA: **Variants at *IRF5-TNPO3*, *17q12-21* and *MMEL1* are associated with primary biliary cirrhosis.** *Nat Genet* 2010, **42**:655-657.
13. Liu X, Invernizzi P, Lu Y, Kosoy R, Bianchi I, Podda M, Xu C, Xie G, Macciardi F, Selmi C, Lupoli S, Shigeta R, Ransom M, Lleo A, Lee AT, Mason AL, Myers RP, Peltekian KM, Ghent CN, Bernuzzi F, Zuin M, Rosina F, Borghesio E, Floreani A, Lazzari R, Niro G, Andriulli A, Muratori L, Muratori P, Almasio PL, et al: **Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis.** *Nat Genet* 2010, **42**:658-660.



14. Sigurdsson S, Padyukov L, Kurreeman FA, Liljedahl U, Wiman AC, Alfredsson L, Toes R, Rönnelid J, Klareskog L, Huizinga TW, Alm G, Syvänen AC, Rönnblom L: **Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis.** *Arthritis Rheum* 2007, **56**:2202-2210.
15. Dieguez-Gonzalez R, Calaza M, Perez-Pampin E, de la Serna AR, Fernandez-Gutierrez B, Castañeda S, Largo R, Joven B, Narvaez J, Navarro F, Marenco JL, Vicario JL, Blanco FJ, Fernandez-Lopez JC, Caliz R, Collado-Escobar MD, Carreño L, Lopez-Longo J, Cañete JD, Gomez-Reino JJ, Gonzalez A: **Association of interferon regulatory factor 5 haplotypes, similar to that found in systemic lupus erythematosus, in a large subgroup of patients with rheumatoid arthritis.** *Arthritis Rheum* 2008, **58**:1264-1274.
16. Dawidowicz K, Allanore Y, Guedj M, Pierlot C, Bombardieri S, Balsa A, Westhovens R, Barrera P, Alves H, Teixeira VH, Petit-Teixeira E, van de Putte L, van Riel P, Prum B, Bardin T, Meyer O, Cornélis F, Dieudé P, ECRAF: **The interferon regulatory factor 5 gene confers susceptibility to rheumatoid arthritis and influences its erosive phenotype.** *Ann Rheum Dis* 2011, **70**:117-121.
17. Wieczorek S, Holle JU, Müller S, Fricke H, Gross WL, Epplen JT: **A functionally relevant IRF5 haplotype is associated with reduced risk to Wegener's granulomatosis.** *J Mol Med* 2010, **88**:413-421.
18. Kristjansdottir G, Sandling JK, Bonetti A, Roos IM, Milani L, Wang C, Gustafsdottir SM, Sigurdsson S, Lundmark A, Tienari PJ, Koivisto K, Elovaara I, Pirttilä T, Reunanen M, Peltonen L, Saarela J, Hillert J, Olsson T, Landegren U, Alcina A, Fernández O, Leyva L, Guerrero M, Lucas M, Izquierdo G, Matesanz F, Syvänen AC: **Interferon regulatory factor 5 (IRF5) gene variants are associated with multiple sclerosis in three distinct populations.** *J Med Genet* 2008, **45**:362-369.
19. Dideberg V, Kristjansdottir G, Milani L, Libioulle C, Sigurdsson S, Louis E, Wiman AC, Vermeire S, Rutgeerts P, Belaiche J, Franchimont D, Van Gossom A, Bours V, Syvänen AC: **An insertion-deletion polymorphism in the interferon regulatory factor 5 (IRF5) gene confers risk of inflammatory bowel diseases.** *Hum Mol Genet* 2007, **16**:3008-3016.
20. Niewold TB, Kelly JA, Flesch MH, Espinoza LR, Harley JB, Crow MK: **Association of the IRF5 risk haplotype with high serum interferon-α activity in systemic lupus erythematosus patients.** *Arthritis Rheum* 2008, **58**:2481-2487.
21. Rullo OJ, Woo JM, Wu H, Hoftman AD, Maranian P, Brahn BA, McCurdy D, Cantor RM, Tsao BP: **Association of IRF5 polymorphisms with activation of the interferon α pathway.** *Ann Rheum Dis* 2010, **69**:611-617.
22. Kozryev SV, Lewén S, Reddy PM, Pons-Estel B, Argentine Collaborative Group, Witte T, German Collaborative Group, Junker P, Lastrup H, Gutiérrez C, Suárez A, Francisca González-Escribano M, Martín J, Spanish Collaborative Group, Alarcón-Riquelme ME: **Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus.** *Arthritis Rheum* 2007, **56**:1234-1241.
23. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, Plenge RM, Koeuth T, Ortmann WA, Hom G, Bauer JW, Gillett C, Burtt N, Cunningham Graham DS, Onofrio R, Petri M, Gunnarsson I, Svenungsson E, Rönnblom L, Nordmark G, Gregersen PK, Moser K, Gaffney PM, Criswell LA, Vyse TJ, Syvänen AC, Bohjanen PR, Daly MJ, Behrens TW, Altschuler D: **Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus.** *Proc Natl Acad Sci USA* 2007, **104**:6758-6763.
24. Sigurdsson S, Göring HH, Kristjansdottir G, Milani L, Nordmark G, Sandling JK, Eloranta ML, Feng D, Sangster-Guity N, Gunnarsson I, Svenungsson E, Sturfelt G, Jönsen A, Truedsson L, Barnes BJ, Alm G, Rönnblom L, Syvänen AC: **Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (IRF5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus.** *Hum Mol Genet* 2008, **17**:872-881.
25. Löfgren SE, Yin H, Delgado-Vega AM, Sanchez E, Lewén S, Pons-Estel BA, Witte T, D'Alfonso S, Ortego-Centeno N, Martín J, Alarcón-Riquelme ME, Kozryev SV: **Promoter insertion/deletion in the IRF5 gene is highly associated with susceptibility to systemic lupus erythematosus in distinct populations, but exerts a modest effect on gene expression in peripheral blood mononuclear cells.** *J Rheumatol* 2010, **37**:574-578.
26. Feng D, Stone RC, Eloranta ML, Sangster-Guity N, Nordmark G, Sigurdsson S, Wang C, Alm G, Syvänen AC, Rönnblom L, Barnes BJ: **Genetic variants and disease-associated factors contribute to enhanced interferon regulatory factor 5 expression in blood cells of patients with systemic lupus erythematosus.** *Arthritis Rheum* 2010, **62**:562-573.
27. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M, Spielman RS: **Natural variation in human gene expression assessed in lymphoblastoid cells.** *Nat Genet* 2003, **33**:422-425.
28. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavaré S, Deloukas P, Dermitzakis ET: **Population genomics of human gene expression.** *Nat Genet* 2007, **39**:1217-1224.
29. Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, Hudson TJ, Sladek R, Majewski J: **Genome-wide analysis of transcript isoform variation in humans.** *Nat Genet* 2008, **40**:225-231.
30. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, Taylor J, Burnett E, Gut I, Farrall M, Lathrop GM, Abecasis GR, Cookson WO: **A genome-wide association study of global gene expression.** *Nat Genet* 2007, **39**:1202-1207.
31. **International HapMap Project.** [http://hapmap.ncbi.nlm.nih.gov/].
32. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: **MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.** *Genet Epidemiol* 2010, **34**:816-834.
33. Davidson R, MacKinnon JG: **Several tests for model specification in the presence of alternative hypotheses.** *Econometrica* 1981, **49**:781-793.
34. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68**:978-989.
35. Kwan T, Benovoy D, Dias C, Gurd S, Serre D, Zuzan H, Clark TA, Schweitzer A, Staples MK, Wang H, Blume JE, Hudson TJ, Sladek R, Majewski J: **Heritability of alternative splicing in the human genome.** *Genome Res* 2007, **17**:1210-1218.
36. **PupSuite 3.1: Prioritization of SNPs in Genotyping Experiments Based on Functional Properties.** [http://pupsuite.bioinfo.cipf.es/].
37. **FastSNP.** [http://fastsnp.ibms.sinica.edu.tw/fastSNP2/].
38. Kelly JA, Kelley JM, Kaufman KM, Kilpatrick J, Bruner GR, Merrill JT, James JA, Frank SG, Reams E, Brown EE, Gibson AW, Marion MC, Langefeld CD, Li QZ, Karp DR, Wakeland EK, Petri M, Ramsey-Goldman R, Reveille JD, Vilá LM, Alarcón GS, Kimberly RP, Harley JB, Edberg JC: **Interferon regulatory factor-5 is genetically associated with systemic lupus erythematosus in African Americans.** *Genes Immun* 2008, **9**:187-194.
39. Kawasaki A, Kyogoku C, Ohashi J, Miyashita R, Hikami K, Kusaoi M, Tokunaga K, Takasaki Y, Hashimoto H, Behrens TW, Tsuchiya N: **Association of IRF5 polymorphisms with systemic lupus erythematosus in a Japanese population: support for a crucial role of intron 1 polymorphisms.** *Arthritis Rheum* 2008, **58**:826-834.

doi:10.1186/ar3343

**Cite this article as:** Alonso-Perez et al.: *Cis-regulation of IRF5 expression is unable to fully account for systemic lupus erythematosus association: analysis of multiple experiments with lymphoblastoid cell lines.* *Arthritis Research & Therapy* 2011 **13**:R80.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit



**Supplementary Table 1:** List of the 109 polymorphisms analyzed in the IRF5 LD block. Polymorphism names in bold indicate the 35 polymorphisms that were selected for the analyses of correlation with IRF5 expression because they were not redundant at a pairwise  $r^2 = 0.90$ . The + sign means that genotypes for these polymorphisms were available in the indicated LCL collection (CEU = European HapMap; Asthma = children with asthma from Dixon et al. [30]) or in the healthy Spanish controls used to complete the LD map (Controls)

Polymorphism	CEU	Asthma	Controls
<b>rs4731528</b>	-	-	+
<b>rs10081379</b>	-	-	+
<b>rs10280295</b>	-	-	+
rs6970780	-	-	+
<b>rs6951243</b>	-	-	+
<b>rs960633</b>	-	-	+
<b>rs4731530</b>	-	-	+
rs4731531	-	-	+
<b>rs2402941</b>	-	-	+
<b>rs6968225</b>	-	-	+
rs6968508	-	-	+
rs6950728	-	-	+
rs11982901	-	-	+
rs4728141	-	-	+
rs13245639	-	-	+
<b>rs729302</b>	+	+	+
rs729068	-	-	+
rs12706860	-	-	+
rs7808659	-	-	+
rs2402940	+	-	+
rs754284	-	-	+

Polymorphism	CEU	Asthma	Controls
rs754281	-	-	+
rs11768806	-	-	+
<b>rs4728142</b>	+	+	+
rs7801838	-	-	+
<b>rs1874330</b>	-	-	+
rs3778754	-	-	+
rs3757388	-	-	+
rs3757387	-	-	+
rs3757385	-	-	+
<b>rs3807134</b>	-	-	+
<b>rs3807135</b>	-	-	+
<b>CGGGG indel</b>	-	-	+
<b>rs2004640</b>	-	-	+
rs3807307	-	-	+
rs752637	+	-	+
rs3823536	-	-	+
rs3778753	-	-	+
<b>rs3807306</b>	+	-	-
<b>rs11761199</b>	+	-	-
<b>rs7808907</b>	+	+	-
<b>rs1874328</b>	+	+	-
<b>In/Del Exon6</b>	-	-	+
<b>rs10954213</b>	-	-	+
rs13242262	+	-	+
<b>rs10488630</b>	+	+	+
<b>rs10488631</b>	+	+	+
<b>rs2280714</b>	+	+	+



Polymorphism	CEU	Asthma	Controls
rs10236569	+	-	-
<b>rs6966125</b>	+	-	-
rs10229001	+	-	-
<b>rs1495458</b>	+	-	-
rs2172876	+	-	-
<b>rs6957529</b>	+	-	-
rs7385716	+	-	-
rs4731535	+	-	+
rs8043	+	-	-
rs1874332	+	+	-
rs2293492	+	+	-
rs12531711	+	-	-
rs17338998	+	-	-
rs2272347	+	-	-
<b>rs3817555</b>	+	-	-
rs7789423	+	+	-
rs6948928	+	+	-
<b>rs12534421</b>	+	-	-
rs12535158	+	-	-
rs12669885	+	-	-
<b>rs1154330</b>	+	+	-
rs17339221	+	-	-
rs2290231	+	-	-
rs11770317	+	-	-
rs1154329	+	-	-
rs6969930	+	-	-
rs2305323	+	-	-

Polymorphism	CEU	Asthma	Controls
rs3958094	+	-	-
rs7807018	+	-	-
<b>rs2305324</b>	+	+	-
rs11768572	+	-	-
rs2305325	+	+	-
rs6965542	+	-	-
rs3847099	+	-	-
rs3857852	+	+	-
rs12539476	+	-	-
<b>rs17424179</b>	+	-	-
rs12155080	+	+	-
rs13236009	+	-	-
rs13221560	+	-	-
rs10239340	+	+	-
rs11762968	+	-	-
rs9649520	+	-	-
rs1072767	+	-	-
rs921403	+	-	-
rs4731541	+	-	-
<b>rs3993439</b>	+	-	-
<b>rs1839600</b>	+	-	-
rs3807301	+	-	-
rs10279821	+	+	-
rs10156169	+	-	-
rs11767238	+	-	-
rs17424602	+	-	-
rs2167273	+	+	-

Polymorphism	CEU	Asthma	Controls
<b>rs6960994</b>	+	+	-
rs6961014	+	+	-
rs6980198	+	-	-
rs2242028	+	-	-
rs13239597	+	-	-
rs11767954	+	-	-
rs13246321	+	-	-

**Supplementary Table 2:** Primers and probes that were used for genotyping the *IRF5* polymorphisms in the healthy Spanish controls by minisequencing. Design of oligonucleotides and protocol of genotyping have already been described (Ferreiro-Neira et al.[4]). Some probes (\$) were extended with a 5' tail (in capitals) that has no homology with human sequences. Others were mutated (underlined capitals) to avoid dimer formation.

SNPs	Oligonucleotides
rs4731530 C>T	Forward cccctcacagtccagtagga Reverse gcccggtggaagtaaattgtc Probe atatttagacattgaaggccccagttct
rs4731531 G>A	Forward taacagatccgcaatggcta Reverse ttctctggggaccatcatag Probe TTACCGagcttggtgcctcacaactctaggggg\$
rs10280295 C>T	Forward ctcaccactccattgcttcc Reverse tgccaagatcatatggcaag Probe gaggatttctggcaatgtccagtgtcattgat
rs6951615 C>T	Forward ggacactgctggttccctta Reverse ggcataaagcacagagtga Probe TTACCTATGATTGATCGTGGTGATATcgccctcagcatcatatgTtgctactgc\$
rs6968508 C>G	Forward ccaatgcaaggggtagtta Reverse ccactatgccagcctaaact Probe tggtatgtgacttatatctcaataaagctgttacaaaaaggggagcg
rs10081379 T>A	Forward acactgacctccagccaagt Reverse cagcaatattcaggggctct Probe TTACCTATGATTGATtactaggtgttctccctagaaatgttatcgatgagaagtcc\$
rs12536195 G>A	Forward acctgtctgcatgctcct Reverse gctatctgtctggcgtctcc Probe TTACCTATGACcggtaagtgccagcctcctcgagtgggt\$
rs11773414 G>A	Forward gtggtgcaatcacagctcac Reverse accgtcctggacaacatagg Probe tcatgcctggctagttttctttattttttaattttgtagagac
rs6955705 T>C	Forward acctcagtctgctcccacag Reverse tcaggagtgagtgggtaggc Probe TTAcagcctcagccccatctcctgcatg\$
rs6970780 T>C	Forward catctccctctgaggtccaa Reverse ccctgtctcattcaccactg Probe agtgccttgcaaataacaccctttcctc
rs6950728 G>A	Forward tggctctgatgtagttgcat Reverse agaaaactgggtggggtgta Probe TTACCTATGAtatgtttataattatcacatcttcctgatggattacacttttatcatt\$
rs4731528 C>T	Forward acctcagtctgctcccacag Reverse tcaggagtgagtgggtaggc Probe TTACCTATGATTGAgcctgccacctgagagtcagcgag\$
rs6951243 G>A	Forward caaggcacttgagttccaca Reverse gtgctcatgtggttcccttt

SNPs	Oligonucleotides
rs6968225 C>G	Probe TTACCTATGATTGATCGTGGgttctgggcaggaccctgtctcattcac\$ Forward gacaacagcctgtcctcaca Reverse ggcctctaactacccccttg Probe TTACCTATGATTGATCGTGGTGATAccataagtcccaccccactgttttagagg\$
rs11768806 C>T	Forward gaggcttgagatggatctgg Reverse cctccggtatgcacttttgt Probe TTACCTATGATTGATCGTGGggtggaTctcgtgggcctggctgggaca\$
rs3757388 A>G	Forward aaaaaggttcccattttgtgg Reverse gctaaggcaggcagatcact Probe TTACCTATGATTGATCGTGGTGATattagcctggcgtgTtgacacacacctatagttc\$
rs2402941 G>A	Forward tgtacctgccccataccttc Reverse gtgtgcccatccaaataacc Probe TTACCTATGATTGATCGTGGTGcaagtggcttctcggcacactgagaa\$
rs754280 G>A	Forward cggaggagatgttcctgaat Reverse gggggcagggttcttactagg Probe TTACCTATGActggccttgagaattgcaaagcaccaaggctcc\$
rs3757387 T>C	Forward gcctcccaagtaggtggaac Reverse atggatggggaaaatgtgaa Probe ccacgccaggctaattttttgtattttttgtagagacaaggttt
rs7801838 C>T	Forward aaagctctgagccggtgtta Reverse gctttgaagtttctggcaca Probe TTACCTATGATTGATggattaatgccaggggcgccacagctgg\$
rs3807134 T>C	Forward gaagctatttgcaccctgga Reverse tctgaaccgttttcgattcc Probe TTACCTATGATTGATCGTGGgagaaggaacaggaggtgtgtgaaggtggaggt\$
rs3807135 C>T	Forward cacatctggaaggggtgtct Reverse tagactggccactggctctt Probe TTACCTATGATTGATCGTGGTGAActccagTccctcctctcgccctgcct\$
rs41298401 C>G	Forward gcgggatgaagactggagta Reverse ggaggcgcttttgaagtc Probe TTACCTATGATTGATCGTccgTgcgccacctgctgtag\$
rs3807307 C>T	Forward cccagaatggggataagtga Reverse ccccaaagtagggccttttag Probe TTACCTATGATTGATCGTGGTagaaccagagagggctcggctg\$
rs3823536 A>G	Forward gtgggtttgcaaggagacat Reverse ctggagtcccaggagacagt Probe TTACCTATGATTGATCGTGGTGATAggacTcactaggggaggaagtgc\$
rs3778753 G>A	Forward ttgattggggtggtctgaat Reverse tccatacaggcagcttaggg Probe TTACCTATGATTGATCGTGGTGATATCCGcaTgagctctaaccggaacagcatccaaac tcc\$
rs11767834 C>T	Forward gtctcaagtgaagggccaaag Reverse ggaaacagaagccacagctc Probe AATGAcagtgaagctAtggcctgggagg\$
rs13245639 C>T <sup>a</sup>	Forward ttggctcattgcaaactctg Reverse ttctcataggaggccagggtg Forward gactacaggcaccaccatc

SNPs	Oligonucleotides
	Reverse tcacacctgtaatcccagca Probe TTACCCacttttgggaggccaaggcggacagatc\$
rs4728141 C>T	Forward tgccaccacgatggttagata Reverse ccaggagttcaaggttacgg Probe TTACCTATGATTGATgatgaaccacagcactccagcctggaca\$
rs960633 G>T	Forward gggctcccgatgttacacta Reverse tggagcgagtggataagggtt Probe CCAGAccctggggggcctcaacattccttgctg\$
rs11982901 C>T	Forward caggtcattcaagggtgtt Reverse taccatcgtggtggcaca Probe TTACCTATGAatcacgccactgcactccagcctgggtg\$
rs1983607 C>T	Forward ggatgcatatttgggttttagcttt Reverse gctgtttacaggaggcacact Probe TTACCctgagtagctgggattacagTtgtgtgccacca\$
rs729068 C>T	Forward caggacttgactggggtat Reverse cctcctgtctcttgccata Probe TTACCTATGATTGATCGTGGgtcagagaatgcaTacgcacagggtgtg\$
rs12706860 C>G	Forward ctcacctggagccacctagt Reverse tccaagaagcatccttcacc Probe ctgtctccaggggtttggtgatgtccag
rs754284 C>G	Forward ggggtgggtctatggatctc Reverse gcctgacacacactgtggaa Probe TTACCTATGATTGATaaccacatggctccctggctctgcctg\$
rs2402940 C>T	Forward tttgcttaggtacaaaggaaaatac Reverse tttaagacgagcctgggtcaa Probe TTACCTATGATTGATCGTGGTGATAcctacttggaggctgaggcaggagaatcactt
rs3778754 G>C	Forward gtgctgctctctgtccttcc Reverse gcacctggctttgctgtatt Probe TTACCTATGATTGATagtaggacctcaaattctcatttcctcgctctctaaaagg\$
rs7808659 C>A	Forward agggagggagagtggagaga Reverse cctcatgatcctcctgcttc Probe tgagccactgtacccggccctaagtaat
rs11763323 G>T	Forward taggagtttcggcaaagggtg Reverse caggcccaagagaaatcaag ProbeTTACCTATGATTGATCGTGGTGATATCCGagaaccgttcccatacatttcatgcc gtccctta\$
rs3757385 G>T	Forward aggggaagtcaaggcagact Reverse tctgaaccgttttcgattcc Probe TTACCcaccttcacacacctcctgttccttctc\$
rs754281 C>T	Forward tgctccaggagagaaaggag Reverse gggggcaggttcttactagg Probe TTACCTATGACctcgagacaggcatgggtcggtggaga\$
rs1874330 T>C	Forward atattccagccaggggaaat Reverse aagggaagtagaggggaag Probe TTACCTATGATTGATCGTGGTGATAgaggcggtgggTacagggaggtctgtcc\$

<sup>a</sup> This SNP was amplified with a nested PCR

**Supplementary Table 3:** Primers that were used for genotyping the indicated polymorphisms in the healthy Spanish controls by Sanger sequencing.

Polimorfismo	Oligonucleotides
rs3778752 T>G and rs3778751 T>A	Forward tgattggggtggtctgaatta Reverse cttttgcctgcagctaggtc
CGGGG indel	Forward cgccgtctggcatctccct Reverse tgagctctgcccaggctgc

**Supplementary Table 1:** Parameters of the multiple linear regression models including known functional polymorphisms to account for IRF5 expression in each of the experimental data sets. Models including exon 6 indel are not shown because there is not any suggestion it could play a role in IRF5 expression. Parameters indicating the fit of the model and the *P* values corresponding to the independent contribution of each polymorphism to models are shown. Expression data sets are labeled as in Table 2.

Data set	Polymorphisms in model	Adjusted $r^2$	Model <i>P</i>	Polymorphism's <i>P</i> -value in model
K8	rs10954213	0.21	$2.3 \times 10^{-4}$	
	rs2004640	0.14	0.0023	
	CGGGG indel	0.13	0.0034	
	rs10954213	0.2	$9.7 \times 10^{-4}$	0.031
	rs2004640			n.s.
	rs10954213	0.19	0.0012	0.027
	CGGGG indel			n.s.
	rs2004640	0.13	0.0098	n.s.
	CGGGG indel			n.s.
S	rs10954213	0.75	$3.3 \times 10^{-19}$	
	rs2004640	0.47	$7.2 \times 10^{-10}$	
	CGGGG indel	0.58	$9.5 \times 10^{-13}$	
	rs10954213	0.76	$1.5 \times 10^{-18}$	$1.4 \times 10^{-11}$
	rs2004640			n.s.
	rs10954213	0.76	$2.2 \times 10^{-18}$	$1.3 \times 10^{-8}$
	CGGGG indel			n.s.
	rs2004640	0.57	$2.4 \times 10^{-11}$	n.s.
	CGGGG indel			$3.5 \times 10^{-4}$
C	rs10954213	0.52	$1.3 \times 10^{-10}$	
	rs2004640	0.33	$1.4 \times 10^{-6}$	
	CGGGG indel	0.36	$3.7 \times 10^{-7}$	
	rs10954213	0.52	$7.7 \times 10^{-10}$	$1.5 \times 10^{-5}$
	rs2004640			n.s.
	rs10954213	0.51	$1.3 \times 10^{-9}$	$9.6 \times 10^{-5}$
	CGGGG indel			n.s.
	rs2004640	0.35	$2.4 \times 10^{-6}$	n.s.
	CGGGG indel			n.s.
D	rs10954213	0.2	$1.3 \times 10^{-10}$	
	rs2004640	0.21	$4.9 \times 10^{-11}$	
	CGGGG indel	0.23	$3.7 \times 10^{-12}$	
	rs10954213	0.27	$1.6 \times 10^{-13}$	$7.2 \times 10^{-5}$
	rs2004640			$2.6 \times 10^{-5}$
	rs10954213	0.27	$4.0 \times 10^{-13}$	0.0028
	CGGGG indel			$6.7 \times 10^{-5}$
	rs2004640	0.23	$1.8 \times 10^{-11}$	n.s.
	CGGGG indel			0.011



**Supplementary Table 2:** Description of the IRF5 haplotypes as defined with tagSNPs in Ferreiro-Neira et al. [4].

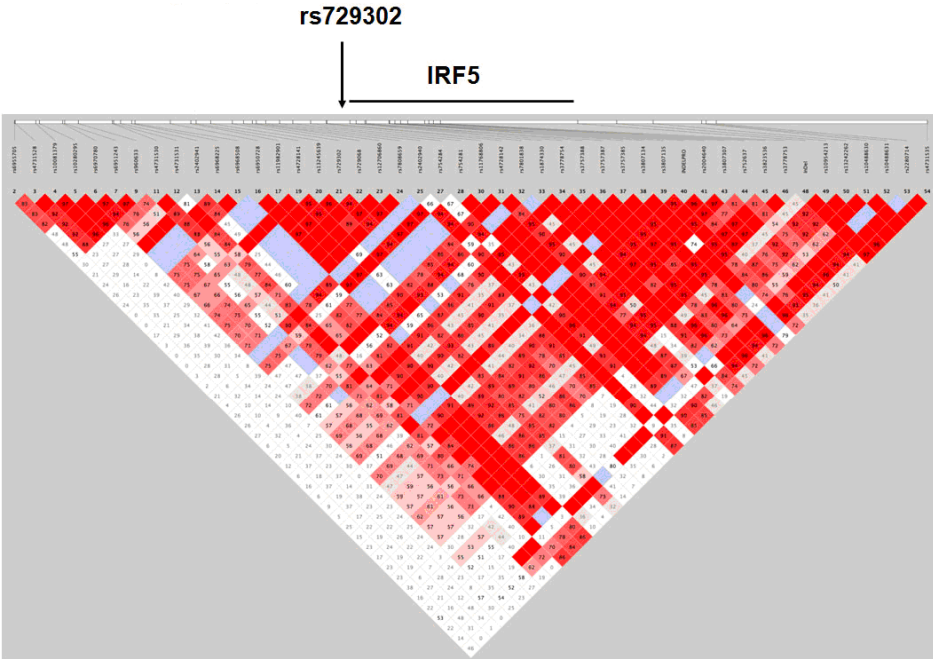
Haplotype	rs729302	rs2004640	rs752637	Exon 6 indel	rs10954213	rs13242262	rs10488630	rs10488631	rs2280714	rs4731535
1	C	G	G	de	A	T	A	T	A	C
2	C	G	A	in	G	A	A	T	G	T
3	A	G	A	in	G	A	A	T	G	T
4	A	T	G	de	A	T	G	T	A	C
5	A	T	G	de	A	T	A	T	A	C
6	A	T	G	in	A	T	A	C	A	T

**Supplementary Table 3:** Lack of correlation between SLE-associated haplotypes defined as in Ferreiro-Neira et al. [4] and the best linear regression model for each of the IRF5 expression data sets. Spearman’s rank order  $r_s^2$  values are given for the carriage of protective haplotypes H1 and H2 and for the risk haplotype H6 taken individually and for the diplotypes made of protector, neutral and risk haplotypes.

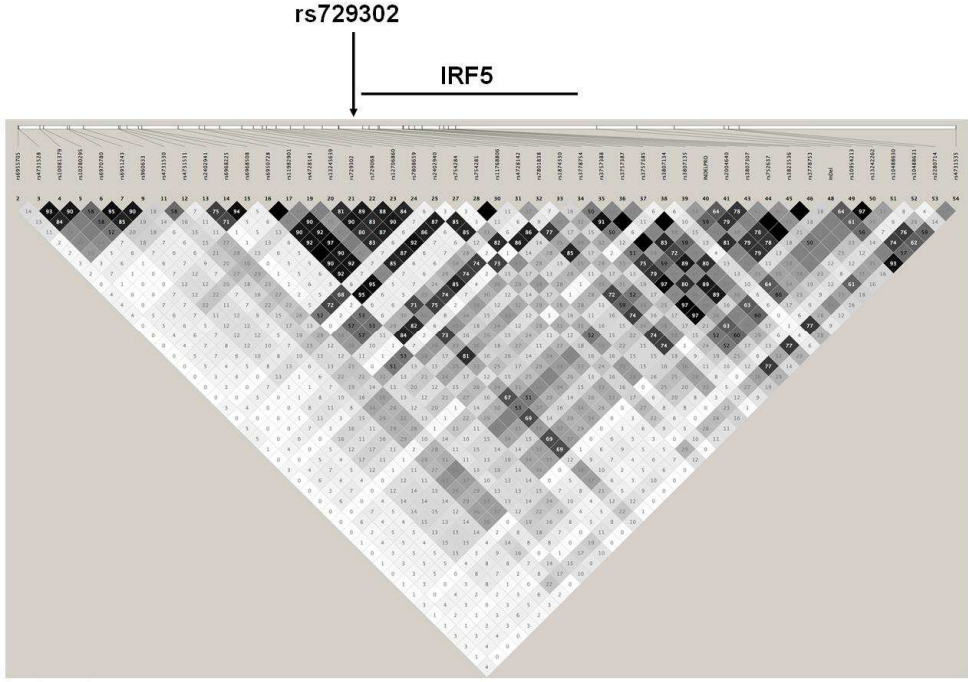
Data set	H1	H2	H6	Diotypes
K8	0.0001	0.04	0.02	0.0005
S	0.08	0.03	0.04	0.003
C	0.008	0.03	0.03	0.12
D	0.15	0.06	0.02	0.06

**Supplementary Figure 1:** Maps of LD in the *IRF5* block representing pairwise A)  $D'$  values and B)  $r^2$  values. Positions of the *IRF5* gene and of rs729302 are highlighted.

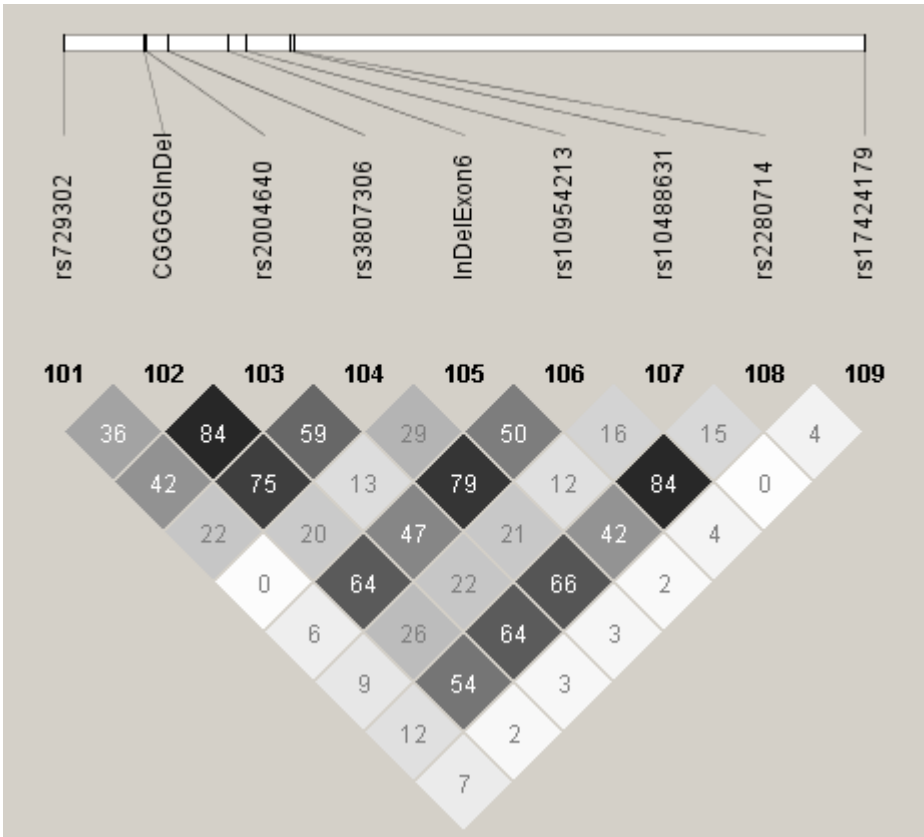
**A.**



**B.**



**Supplementary Figure 2:** Pairwise  $r^2$  between the most relevant polymorphisms in our analysis (detail from Supplementary Figure 1B).



**Supplementary Information:**

**Expression of the IRF5 isoforms:** Some of the studies in our analysis had included probes targeting IRF5 isoforms (Figure 1 in the main text). Expression results from the probe targeting exon 1C did not show association with any of the IRF5 polymorphisms, suggesting that its alternative splicing is not cis-regulated. This was in contrast with a recent report showing association of rs10488631 with expression of this exon in SLE blood cells [26]. Two of the studies contained a probe targeting the 3'UTR posterior to rs10954213. Presence of this part of the 3'UTR was largely dependent on this SNP (with  $P$  values  $9.7 \times 10^{-21}$  in Kwan's data and  $1.4 \times 10^{-33}$  in Dixon's data) as has already been described [3,23]. No meaningful analysis of expression data for exon 1A was possible given the widely discordant results obtained in the two studies targeting this exon (Figure 2 in the main text). Also, no analysis was done for the probe in variant 5 of exon 6 given its very low expression level.

**Mapa de LD de IRF5: búsqueda de  
polimorfismos funcionales.**

### 1. Pacientes y controles.

Para la realización del mapa de LD de la región promotora del gen *IRF5* se han utilizado 95 muestras control procedentes del Hospital Clínico Universitario de Santiago de Compostela, los cuales no presentaban historia de padecer ningún tipo de enfermedad autoinmune.

### 2. Extracción de ADN.

El ADN se extrajo a partir de sangre periférica utilizando el protocolo clásico de extracción con fenol-cloroformo.

### 3. Polimorfismos analizados.

Para la realización del mapa de LD se genotiparon un total de 54 polimorfismos incluyendo los 11 polimorfismos del gen *IRF5* previamente estudiados por nuestro grupo (**Publicación 3**) (**Tabla 1**). Todos ellos están dentro de la región chr7:128343695-128393203 (*Human March* 2006; NCBI36/hg18 *Assembly*). Esta región abarca el gen *IRF5* en su totalidad, 21,5 Kb de su región promotora (referenciado al inicio del exón 1 variante A) y parte del gen *TNPO3*.

**Tabla 1: Polimorfismos incluidos en el mapa LD.**

Polimorfismo	Posición	Localización	Alelos	MAF	HapMap	1000 Genomes
rs12536195	128343695	promotor <i>IRF5</i>	C < G	0,00%	-	-
rs6955705	128344606	promotor <i>IRF5</i>	C < T	16,70%	-	+
rs4731528	128344642	promotor <i>IRF5</i>	T < C	49,50%	-	+
rs10081379	128345738	promotor <i>IRF5</i>	T < A	48,90%	-	+
rs10280295	128345925	promotor <i>IRF5</i>	T < C	50,00%	-	+
rs6970780	128347164	promotor <i>IRF5</i>	C < T	34,00%	-	+
rs6951243	128347271	promotor <i>IRF5</i>	A < G	34,00%	-	+
rs6951615	128347598	promotor <i>IRF5</i>	T < C	0,50%	-	-
rs960633	128347997	promotor <i>IRF5</i>	T < G	33,20%	+	+
rs11773414	128349154	promotor <i>IRF5</i>	A < G	0,00%	-	+
rs4731530	128349814	promotor <i>IRF5</i>	T < C	39,90%	-	+
rs2402941	128350250	promotor <i>IRF5</i>	A < G	11,70%	-	+
rs6968225	128350843	promotor <i>IRF5</i>	G < C	11,20%	+	+

Polimorfismo	Posición	Localización	Alelos	MAF	HapMap	1000 Genomes
rs6950728	128352897	promotor <i>IRF5</i>	A < G	30,90%	-	+
rs11982901	128353990	promotor <i>IRF5</i>	T < C	31,60%	-	+
rs11983607	128354093	promotor <i>IRF5</i>	T < C	0,00%	-	-
rs4728141	128354268	promotor <i>IRF5</i>	T < C	49,50%	+	+
rs13245639	128355046	promotor <i>IRF5</i>	T < C	31,60%	-	+
rs729302	128356196	promotor <i>IRF5</i>	C < A	31,10%	+	+
rs729068	128356797	promotor <i>IRF5</i>	T < C	31,10%	-	+
rs12706860	128357262	promotor <i>IRF5</i>	G < C	33,70%	-	+
rs7808659	128357474	promotor <i>IRF5</i>	A < C	30,40%	-	+
rs2402940	128357978	promotor <i>IRF5</i>	T < C	13,20%	+	+
rs11763323	128358369	promotor <i>IRF5</i>	T < G	0,00%	+	-
rs754284	128358714	promotor <i>IRF5</i>	G < C	31,60%	-	+
rs754281	128359416	promotor <i>IRF5</i>	T < C	12,40%	-	+
rs754280	128359539	promotor <i>IRF5</i>	A < G	0,00%	+	-
rs11768806	128360364	promotor <i>IRF5</i>	T < C	12,40%	-	+
rs4728142	128361203	promotor <i>IRF5</i>	A < G	47,80%	+	+
rs7801838	128361230	promotor <i>IRF5</i>	T < C	28,90%	-	+
rs1874330	128362460	promotor <i>IRF5</i>	C < T	14,00%	-	+
rs3778754	128362788	promotor <i>IRF5</i>	C < G	48,90%	-	+
rs3757388	128363259	promotor <i>IRF5</i>	G < A	33,30%	-	+
rs3757387	128363322	promotor <i>IRF5</i>	C < T	49,40%	-	+
rs3757385	128364540	promotor <i>IRF5</i>	T < G	33,30%	-	+
rs3807135	128364853	promotor <i>IRF5</i>	T < C	33,00%	-	+
rs2004640	128365537	intrón 1 <i>IRF5</i>	G < T	44,00%	-	+
rs41298401	128365543	intrón 1 <i>IRF5</i>	G < C	0,60%	-	-
rs3807307	128366438	intrón 1 <i>IRF5</i>	T < C	47,80%	-	+
rs752637	128366656	intrón 1 <i>IRF5</i>	T < C	39,80%	+	+
rs3823536	128366902	intrón 1 <i>IRF5</i>	G < A	47,80%	-	+
rs3778753	128367278	intrón 1 <i>IRF5</i>	A < G	47,80%	-	+
rs3778752	128367283	intrón 1 <i>IRF5</i>	G < T	47,80%	-	+
rs3778751	128367284	intrón 1 <i>IRF5</i>	A < T	47,80%	-	+
rs3807306	128367916	intrón 1 <i>IRF5</i>	G < T	45,70%	+	+
rs11767834	128368512	intrón 1 <i>IRF5</i>	T < C	1,10%	-	-
indel exón 6	128374588	exón 6 <i>IRF5</i>	I < D	45,80%	-	-
rs10954213	128376663	3'UTR <i>IRF5</i>	G < A	35,10%	-	+
rs13242262	128378600	intergénico	A < T	35,00%	+	+
rs10488630	128381184	promotor <i>TNPO3</i>	G < A	40,90%	+	+
rs10488631	128381419	promotor <i>TNPO3</i>	C < T	11,10%	+	+
rs2280714	128381961	exón 1 <i>TNPO3</i>	C < T	27,20%	+	+
rs4731535	128393203	intrón 2 <i>TNPO3</i>	C < T	46,70%	+	+

**Tabla 1: Polimorfismos incluidos en el mapa LD.** MAF: frecuencia del alelo menor en nuestra colección de 95 muestras control de Santiago de Compostela. +/-: presente o no en el proyecto internacional HapMap o 1000 Genomes.



#### 4. Genotipado.

50 SNPs fueron genotipados mediante minisecuenciado con la técnica *SNaPshot*® (Applied Biosystems)\*<sup>1</sup>. El primer paso de la técnica *SNaPshot*® consiste en amplificar mediante PCR multiplex los fragmentos de las secuencias donde se encuentran incluidos los polimorfismos a genotipar. Para llevar a cabo esta PCR multiplex utilizamos el *Multiplex PCR kit* (QUIAGEN). La reacción de PCR se realizó en 10 µL de volumen final conteniendo 20 ng de ADN genómico, 1X *QUIAGEN*® *Multiplex PCR Master Mix* (concentración de MgCl<sub>2</sub> 3mM) y 0,2 µM de cada *primer*. Las condiciones de la PCR multiplex son las siguientes: desnaturalización inicial a 95°C durante 15 minutos, seguida de 30 ciclos de desnaturalización a 94°C durante 30 segundos, hibridación a 60°C durante 90 segundos y una extensión a 72°C durante 90 segundos. Finalmente se realiza una fase de extensión a 72°C durante 10 minutos.

Una vez amplificados los fragmentos se lleva a cabo un tratamiento EXO-SAP con el fin de eliminar dNTPs no incorporados. A continuación, se realiza la reacción de *SNaPshot*® propiamente dicha en la que se utiliza el *SNaPshot*® *Multiplex Kit* (Applied Biosystems). La reacción de *SNaPshot*® se realizó en 5 µL de volumen final conteniendo 1 µL de multiplex PCR purificada, 1 µL de *SNaPshot*® *Multiplex Ready Reaction Mix*, 1,2 µL de betaína y 0,2 µM de cada sonda. Las condiciones de la reacción son las siguientes: 25 ciclos de desnaturalización de 96°C durante 10 segundos, hibridación de 50°C durante 5 segundos y una extensión de 60°C durante 30 segundos. Sobre el producto de *SNaPshot*® se lleva a cabo un segundo tratamiento con SAP para eliminar los ddNTPs no incorporados. Por último, los fragmentos purificados se separan mediante electroforesis capilar en el secuenciador *3130xl Genetic Analyzer* (Applied Biosystems). Los resultados obtenidos se analizaron posteriormente mediante el programa *Gene Mapper*® v4.0. (Applied Biosystems).

Los *primers* utilizados para cada polimorfismo fueron diseñados mediante los programas *Oligos*© 1999-2002 v.9.11 y *Primer 3* ([http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)). Todos los oligonucleótidos fueron sintetizados por *SIGMA-Genosys*. En la **tabla 2** están indicados los 54 polimorfismos genotipados en este estudio junto con los oligonucleótidos utilizados. Algunas sondas contienen colas no humanas (escritas en mayúsculas) que no hibridan con el genoma humano. Estas colas se añaden a las sondas para

\*<sup>1</sup> *SNaPshot*® es una técnica multiplex que permite genotipar varios SNPs en una misma reacción. Su fundamento se basa en la utilización de ddNTPs (dideoxi-nucleótido-trifosfato) marcados con fluoróforos diferentes los cuales detienen la reacción de polimerización al ser incorporados. La reacción de *SNaPshot*® propiamente dicha consiste en la generación de fragmentos por elongación de un único ddNTP marcado que es complementario al nucleótido del SNP. Los oligonucleótidos utilizados como sondas se diseñan de modo que sean complementarios hasta una base antes de la posición donde se encuentra localizado el SNP. De este modo, la única base que se añade corresponde exactamente con el lugar donde se encuentra el SNP.

aumentar su longitud. Por otro lado, ciertas sondas de *SNaPshot*® presentan mutaciones puntuales en su secuencia (representadas en mayúsculas y subrayadas). Estas mutaciones se introducen para evitar la formación de estructuras secundarias.

Los SNPs rs3778752, rs3778751 se genotiparon mediante secuenciación. Concretamente, fueron amplificados en el mismo fragmento de PCR ya que están situados en bases contiguas. Para la amplificación del fragmento se realizó una reacción de PCR de 10 µL de volumen final con 25 ng de ADN, 1X PCR buffer de reacción (ROCHE) con una concentración 1,5 mM de MgCl<sub>2</sub> y 500 mM de KCl (pH = 8,3), 0,2 mM de dNTPs (deoxi-nucleótidos-trifosfato), 1U de *Taq DNA Polymerase* (ROCHE) y 0,1 µM de cada *primer*. Las condiciones de amplificación consistieron en una desnaturalización genómica de 95°C durante 3 minutos, 35 ciclos de desnaturalización a 95°C durante 15 segundos, hibridación a 56°C durante 15 segundos y extensión a 72 °C durante 15 segundos. Para la realización de esta PCR aplicamos un *hot-start* manual. Esta técnica reduce la amplificación inespecífica durante las etapas iniciales de la PCR, es decir, mientras que el termociclador alcanza la temperatura de desnaturalización del ADNdc (95 °C). Para evitar esto, la PCR *hot-start* se basa en que la reacción comience cuando el termociclador ya esté a 95° C.

Seguidamente, la reacción de secuencia se realiza en un volumen final de 5 µL que contiene 2,5 µL de ADN amplificado y previamente tratado con EXO (*Exonuclease I*, Epicentre, Madison, WI) y SAP (*Shrimp Alkaline Phosphatase*, GE Healthcare, Barcelona, España), 0,3 µM de cada uno de los *primers* y 2 µL de *Big Dye*® *Terminator v 3.1. Cycle Sequencing Kit* (Applied Biosystems). Las condiciones de secuenciación fueron: desnaturalización inicial 96°C durante 4 minutos, 30 ciclos de desnaturalización a 96°C durante 15 segundos, hibridación a 50°C durante 10 segundos y extensión a 60°C durante 3 minutos; por último, se hace una extensión final a 60°C durante 10 minutos. A continuación, realizamos la purificación del producto de secuenciación con *BigDye X Terminator*® *Purification Kit* (Applied Biosystems). Finalmente, la lectura de los productos de secuenciación se realiza en el secuenciador *3130xl Genetic Analyzer* (Applied Biosystems) y el análisis de la secuencia mediante el programa *Sequencing Analysis v5.2*.

Para el indel del exón 6 la técnica empleada consistió en una amplificación por PCR seguida de una separación por tamaño de fragmento mediante electroforesis en geles de agarosa al 3%. Para la amplificación del fragmento se realizó una reacción de PCR de 10 µL de volumen final con 25 ng de ADN, 1X PCR buffer de reacción (ROCHE) con una concentración 1,5 mM de MgCl<sub>2</sub> y 500 mM de KCl (pH = 8,3), 0,2 mM de dNTPs, 1U de *Taq DNA Polymerase* (ROCHE) y 0,1 µM de cada *primer*. Las condiciones de amplificación consistieron en una desnaturalización inicial de 95°C durante 3 minutos seguida de 35 ciclos de desnaturalización a

95°C durante 15 segundos, hibridación durante 15 segundos a 60°C y una extensión a 72°C durante 15 segundos. Para la realización de esta PCR aplicamos un *hot-start* manual.

El SNP rs10954213 (señal de poly A) se genotipó mediante la tecnología *TaqMan*<sup>\*2</sup> (*TaqMan MBG Genotyping Assay*, Applied Biosystems) con una sonda complementaria a cada uno de los dos alelos marcada con los fluoróforos FAM ó VIC. La reacción de amplificación se realizó en un volumen final de 10 µL con 25 ng de ADN, 1X *TaqMan Universal PCR Master Mix, no amperase UNG* (Applied Biosystems) y 1X de la mezcla que contiene los *primers* y sondas FRET específicos para el genotipado del SNP rs10954213. Las condiciones de amplificación consistieron en una desnaturalización inicial de 95°C durante 10 minutos seguida de 40 ciclos de desnaturalización a 95°C durante 15 segundos e hibridación durante 1 minuto a 60°C. Para la lectura de los experimentos se utilizó el *Chromo4 real-time PCR system* (MJ Research, Waltham, MA, USA).

Por último, el elevado porcentaje de A (adenina) y T (timina) en las zonas adyacentes al SNP rs13245639 hizo que este polimorfismo se amplificara mediante *nested* (anidada) PCR. La *nested* PCR es una variante de la PCR convencional que comprende dos rondas de amplificación con distintos pares de *primers* en cada una, con el fin de incrementar la sensibilidad de detección. Primero se realiza una reacción con los *primers* externos (denominados como *forward\_1* y *reverse\_1* en la **tabla 2**) para amplificar una región de ADN más extensa, que contiene el polimorfismo. Después, con este producto de amplificación, se realiza una segunda PCR con los *primers* internos (denominados como *forward\_2* y *reverse\_2* en la **tabla 2**) para amplificar la región específica. A partir de esta amplificación se llevó a cabo el mismo protocolo de *SNaPshot*® descrito previamente. La tasa de genotipado de los 54 polimorfismos fue de 98,91 %.

---

<sup>\*2</sup> La tecnología *TaqMan* se basa en la utilización de sondas FRET. Para el genotipado se emplean dos sondas FRET siendo cada una de ellas complementaria a la secuencia de uno de los alelos del SNP. Estas sondas FRET están marcadas en su extremo 5' con un fluoróforo (*reporter*) cuya emisión de fluorescencia es captada por otro fluoróforo (*quencher*) que absorbe la fluorescencia emitida por el *reporter* que está situado en el extremo 3' de la misma sonda. Después de la hibridación y durante la PCR, la actividad 5' exonucleasa de la Taq polimerasa libera al *reporter* que queda en solución y emite fluorescencia al ser excitado.

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs12536195	Forward acctgtctgcgatgctcct Reverse gctatctgtctggcgtctcc Sonda TTACCTATGACcggaagtgccagcctcctcgagtgg
rs6955705	Forward acctcagtctgctccacag Reverse tcaggagtgagtggttaggc Sonda TTAacgctcagccccatctcctcatg
rs4731528	Forward acctcagtctgctccacag Reverse tcaggagtgagtggttaggc Sonda TTACCTATGATTGAgcctgccacctgagagtcagcgag
rs10081379	Forward aactgacctccagccaagt Reverse cagcaatattcagggctct Sonda TTACCTATGATTGATtactaggtgttctccctagaatgtatcgtcatgagaagtcc
rs10280295	Forward ctccactccattgcttcc Reverse tgccaagatcatatggcaag Sonda gaggatttctggcaatgtccagtgtcattgat
rs6970780	Forward catctccctctgaggtcaa Reverse ccctgtctcattcaccactg Sonda agtgccttgcaaataacaccctttctc
rs6951243	Forward caaggcacttgagttccaca Reverse gtgctcatgtgttccttt Sonda TTACCTATGATTGATCGTGGgttctgggcaggaccctgtctcattcac
rs6951615	Forward ggacactgctggttcctta Reverse ggcatgaaagcacagagtga Sonda TTACCTATGATTGATCGTGGTGATATcgctcagcatcatatgTtgctactgc
rs960633	Forward gggctcccgatgttaccta Reverse tggagcgagtggaaggtt Sonda CCAGAcctgggggctcaacattccttgctg
rs11773414	Forward gtggtgcaatcacagctcac Reverse accgtcctggacaacatagg Sonda tcatgctggctagttttcttatttttaattttagagac

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas.

<sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipo mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).

**Continuación tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs4731530	Forward cccctcacagtccagtagga Reverse gccctggaagtaaattgtc Sonda atatttagacattgaagccccagttct
rs2402941	Forward tgtacctgccccataccttc Reverse gtgtgcccatacaataacc Sonda TTACCTATGATTGATCGTGGTGcaagtggcttctcgccacactgagaa
rs6968225	Forward gacaacagcctgtcctcaca Reverse ggcctctaactaccccttg Sonda TTACCTATGATTGATCGTGGTGATAccataagtcaccccccactgttttagagg
rs6968508	Forward ccaatgcaaggggtagtta Reverse ccactatgcccagcctaaact Sonda tggatgtgacttatctcaataaagctgttcaaaaaaggggagcg
rs6950728	Forward tggctctgatgttagttgcat Reverse agaaaactgggtgggctgta Sonda TTACCTATGAtatgtttataattatcacatcttctgatggattacacttttatcatt
rs11982901	Forward caggctattcaagggtggtt Reverse taccatcgtgggtggcaca Sonda TTACCTATGAatcacgccactgcactccagcctgggtg
rs11983607	Forward ggatgcatatttggtttagcttt Reverse gctgtttacaggaggcacact Sonda TTACCctgagtagctgggattacagTtgtgtgccacca
rs4728141	Forward tgccaccacgatggtagata Reverse ccaggagttcaaggttacgg Sonda TTACCTATGATTGATgatgaaccacagcactccagcctggaca
rs13245639 <sup>a</sup>	Forward_1 ttggctcattgcaaactctg Reverse_1 ttctcataggaggccaggtg Forward_2 gactacaggcaccaccatc Reverse_2 tcacacctgtaatcccagca Sonda TTACCcactttgggaggccaaggcggacagatc
rs729302	Forward acccactctgcacatctg Reverse cccaaccaagggtgaacc Sonda ctgggtggactctggtgtgtagtgatcctg

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas.

<sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipo mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).

**Continuación tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs729068	Forward caggacttgactggggtat Reverse cctcgtgtctcttgccata Sonda TTACCTATGATTGATCGTGGgtcagagaatgcaT <sup>a</sup> acgcacaggtgtg
rs12706860	Forward ctacactggagccacctagt Reverse tccaagaagcatccttcacc Sonda ctgtctccaggggttggatgtccag
rs7808659	Forward agggaggagagtgagagaga Reverse cctcatgatcctctgcttc Sonda tgagccactgtaccggccctaagtaat
rs2402940	Forward ttgcttaggtacaaaggaaaatac Reverse ttaagacgagcctgtgtaa Sonda TTACCTATGATTGATCGTGGTGATAcctacttgaaggctgaggcaggagaatcactt
rs11763323	Forward taggagtttcggcaaaggtg Reverse caggccaagagaaatcaag Sonda TTACCTATGATTGATCGTGGTGATATCCGagaaccgttcccatcacttcgctgcctta
rs754284	Forward ggggtgggtctatggatctc Reverse gcctgacacacactgtggaa Sonda TTACCTATGATTGATaaccacatggctccctggctctgcctg
rs754281	Forward tgctccaggagagaaaggag Reverse gggggcaggttcttactagg Sonda TTACCTATGAcctcgagacaggcatgggtcgggtggaga
rs754280	Forward cggtggagatgttcctgaat Reverse gggggcaggttcttactagg Sonda TTACCTATGActggttgagaattgcaagcaccaaggctcc
rs11768806	Forward gaggcttgagatggatctgg Reverse cctccggtatgcacttttgt Sonda TTACCTATGATTGATCGTGGgtggaT <sup>a</sup> ctcgtgggcctggctgggaca
rs4728142	Forward cgggcagcttctccacagg Reverse agccgagattgcacattgc Sonda GGTTCGTCTCTCGTTGACATGAGGcacagctgggccttctccccatttctactaacac

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.** Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas. <sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipó mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).

**Continuación tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs7801838	Forward aaagctctgagccggtgta Reverse gctttgaagtttctggcaca Sonda TTACCTATGATTGATggattaatgccaggcgccacagctgg
rs1874330	Forward atattccagccagggaat Reverse aagggaagtagagggaag Sonda TTACCTATGATTGATCGTGGTGATAgaggcggtgggTacagggaggtctgtcc
rs3778754	Forward gtgctgctctctgctctcc Reverse gcacctggctttgctgtatt Sonda TTACCTATGATTGATagtaggacctcaaatctatttctcgtctctaaaagg
rs3757388	Forward aaaaaggttcccattttgtgg Reverse gctaaggcagcgagatcact Sonda TTACCTATGATTGATCGTGGTGATAttgacctggcgtgTgacacacacctatagttc
rs3757387	Forward gcctccaagtaggtggaac Reverse atggatggggaaaaatgtgaa Sonda ccacgccaggctaattttgtattttttagagacaaggttt
rs3757385	Forward aggggaagtcaaggcagact Reverse tctgaaccgttttctgattcc Sonda TTACCaccttcacacacctctgttcctctc
rs3807135	Forward cacatctggaagggtgtct Reverse tagactggccactggctctt Sonda TTACCTATGATTGATCGTGGTGActccagTccctcctctcgctgcct
rs2004640	Forward cgctactgcccccaagtcta Reverse gagggagagcagcagagc Sonda GGTTCTGTTCTTCGTTGACATGAGGgccggccggcgccacctgctgtaggcaccc
rs41298401	Forward gcgggatgaagactggagta Reverse ggaggcgctttggaagtc Sonda TTACCTATGATTGATCGTccgTgcgcacctgctgtag
rs3807307	Forward cccagaatggggataagtga Reverse ccccaaagttagggcctttag Sonda TTACCTATGATTGATCGTGGTagaaccagagagggtcgcgtg

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas. <sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipo mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).

**Continuación tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs752637	Forward tgcccacctggtccagacag Reverse ggagccctctggcaagacc Sonda accctgacctgggaggaagc
rs3823536	Forward gtgggtttgcaaggagacat Reverse ctggagtcccaggagacagt Sonda TTACCTATGATTGATCGTGGTGATAGgacTcactaggggaggaagtgc
rs3778753	Forward ttgattgggtggtctgaat Reverse tccatacaggcagcttaggg Sonda TTACCTATGATTGATCGTGGTGATATCCGcaTgagctctaaccgaacagcatccaaactcc
rs3778752 <sup>b</sup> rs3778751 <sup>b</sup>	Forward tgattgggtggtctgaatta Reverse ctttgcctgcagctaggtc
rs3807306	Forward gccacagtgagtctggtt Reverse caccaaccaggagaggtaa Sonda TTACCTATGATTGATCGTGGTGATATCCGggccccagggaacgaaagtggctagac
rs11767834	Forward gtctcaagtgaaggccaag Reverse ggaaacagaagccacagctc Sonda AATGAcagtgaagctAtggcctgggagg
indel exón 6 <sup>c</sup>	Forward gaggtggcactgacagccgtc Reverse ggggaccagcaccacgg
rs10954213 <sup>d</sup>	Forward gtccaagaacctggagcagaaataa Reverse ggagcaactggtagtagtaaaagga TagMan Sonda tctgttttaacatt[c/t]attaat
rs13242262	Forward gatgtcacatatctggttccagc Reverse tggggaagtcagccactcg Sonda tggcaaatatgaaaacactgtcatattcctcaaaggcaa
rs10488630	Forward tgataacggtttcatggatgt Reverse tcatcagccttacattttacc Sonda GGTCTGTTCTTCGTTGACATGAGGggtcaggaaactgtctactgggtggccatagtagaaaaacc
rs10488631	Forward tgtaagcgtgatgagctttgg Reverse cagcccttcacagccatcag Sonda tgttctttgtttttactttcattcgtcctgtagctcggaatggttc

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas.

<sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipo mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).



**Continuación tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Polimorfismo	Oligonucleótidos
rs2280714	Forward caaggaggagtaagcaaggaac Reverse cccagtttggctcttctc Sonda ggtcacattggtgatgctc
rs4731535	Forward attctggcacagctacaac Reverse tgttggcacaaccaatctca Sonda gtatctagtagtcagactcaacattaagtacgaagtaattgc

**Tabla 2: Oligonucleótidos empleados para el genotipado de los 54 polimorfismos del mapa de LD.**

Todas las secuencias de los oligonucleótidos están escritas en sentido 5'-3'. Las colas no humanas están representadas en mayúsculas. Las mutaciones puntuales están representadas en mayúsculas y subrayadas. <sup>a</sup> Este SNP rs13245639 se amplificó mediante *nested* PCR. <sup>b</sup> Estos polimorfismos fueron genotipados mediante secuenciación. <sup>c</sup> El indel del exón 6 se genotipó con una PCR convencional seguida de una separación por tamaño de fragmento mediante electroforesis. <sup>d</sup> El SNP rs10954213 se genotipo mediante la tecnología *TaqMan* (*TaqMan MBG Genotyping Assay*, Applied Biosystems).

## 5. Cultivo celular.

Para la realización de los ensayos funcionales *in vitro*, se ha utilizado una línea linfocitaria B llamada WIL2 NS (ECACC n°: 90112121) obtenida de la Colección Europea de Cultivos Celulares. Se cultivaron en medio RPMI 1640 (SIGMA) completo, es decir, suplementado con 10% de suero fetal bovino (FBS) (SIGMA), 2 mM de L-glutamina (SIGMA), 1 U/mL de penicilina y 0,1 mg/mL de estreptomycin (SIGMA). Las células en cultivo se encuentran en un ambiente húmedo a 37° C y con un 5% de CO<sub>2</sub>.

Cuando se indica esta línea celular se estimuló durante 4 horas mediante la adición al cultivo de 500 ng/mL de ionomicina y 10 ng/mL de PMA (*phorbol 12-myristate 13-acetate*).

## 6. Extracción de proteínas nucleares.

Para la obtención de los extractos proteicos nucleares a partir de la línea celular WIL 2 NS se utilizó el siguiente protocolo:

- Se lavan 10<sup>7</sup> células con 1 mL de buffer fosfato salino (PBS) 1X a 4° C y se recogen centrifugando a 4° C durante 5 minutos a 4000 rpm. El pellet se resuspende en 200 µL de buffer

A (10 mM HEPES pH:7,9, 1 mM EDTA, 1 mM EGTA, 10 mM KCl, 1 mM DTT, 1 mM PMSF, 10 µg/mL de aprotinina y 10 µg/mL de leupeptina).

- Las células se lisan durante una incubación en hielo de 15 minutos.
- Añadimos 0,1% de detergente no iónico (Triton X-100) e incubamos 10 minutos en hielo. Agitamos vigorosamente durante 15 segundos. Centrifugamos a 4°C durante 15 minutos a 800 xg. Retiramos el sobrenadante (fracción citosólica) y resuspendemos el *pellet* (núcleos celulares) en el buffer A añadiendo 20% de glicerol y 0,4 M de KCl.
- Incubamos la muestra a 4° C en agitación suave durante 30 minutos. Centrifugamos a 1300 rpm y 4° C durante 15 minutos precipitando así el material insoluble.
- Por último, se transfiere el sobrenadante (extracto proteico nuclear) a un nuevo tubo para cuantificar la proteína utilizando *Quant-iT™ Protein Assay Kit* (Invitrogen) y leyendo en el *Qubit™ fluorometer* (Invitrogen). Los extractos se alicuotaron y almacenaron a - 80° C.

## 7. Ensayo de cambio de movilidad electroforética (EMSA, *Electroforetic Mobility Shift Assay*).

Las interacciones entre proteínas y ADN tienen un papel central en el control de muchos procesos celulares como: replicación, recombinación y reparación del ADN, transcripción y ensamblaje viral. Esta técnica es imprescindible para estudiar la regulación génica y determinar las interacciones proteína-ADN. El EMSA se basa en que los complejos proteína-ADN migran más lentamente que los oligonucleótidos libres de doble cadena en una electroforesis. Así, el ensayo se lleva a cabo incubando extractos proteicos nucleares con los oligonucleótidos de doble cadena marcados. Los productos de la reacción se analizan mediante una electroforesis en gel de poliacrilamida no desnaturizante. La especificidad de la unión se establece mediante experimentos de competición, es decir, pre-incubando con un exceso del mismo oligonucleótido de doble cadena no marcado.

Para la realización de esta técnica, se utilizaron oligonucleótidos sintéticos marcados con biotina para los dos alelos del polimorfismo a estudiar. Los oligonucleótidos sintéticos utilizados se muestran en la **tabla 3**. Todos ellos fueron sintetizados por SIGMA-Genosys. El protocolo que se siguió es el del *Lighthift™ Chemiluminiscent EMSA kit* (PIERCE) con algunas modificaciones. En un primer paso, se hibridan las secuencias complementarias de los

oligonucleótidos de cadena sencilla durante 1 hora a 37°C. Después, se realizan las reacciones de *binding* con 0,5 - 6 µg de extracto proteico nuclear de la línea WIL2 NS y 0,5 nM de los oligonucleótidos de doble cadena marcados e hibridados. Esta reacción se realiza en un volumen final de 20 µL con las siguientes condiciones: 10 mM Tris\_HCL pH: 7,5, 10 mM KCl, 0,6 mM DTT, 50 ng/µL poly(dI•dC), 1,5 mM EDTA, 2 mM HEPES pH: 7,9, 25 mM NaCl, 0,1 mM ZnSO<sub>4</sub>, 15% glicerol y 0,25 mg/mL BSA. Esta reacción se incuba a temperatura ambiente durante 20 minutos. Los controles incluyen la reacción de competición, con un exceso del oligonucleótido de doble cadena no marcado (200X). A continuación, las muestras se cargan en un gel de poliacrilamida 5% no desnaturalizante. La electroforesis se realiza a 11 V/cm, 0,5% TBE y en hielo. Los complejos proteína-ADN se traspasan a una membrana de nylon mediante una transferencia semiseca (45 minutos a 12V). Después, se hace un *cross-linking* del ADN a la membrana a 7,2 J/cm<sup>2</sup> en el transiluminador *Fluo-Link* (Vilber Lourmat).

El revelado consiste en incubar la membrana en las siguientes soluciones, a temperatura ambiente y en agitación suave:

- 20 mL de *Lightshift Blocking Buffer* durante 15 minutos para bloquear la membrana.
- 20 mL de *Lightshift Blocking Buffer* con 67,7 µL de conjugado estreptavidina-peroxidasa durante otros 15 minutos. Con esta incubación, se permite que la estreptavidina se una a los oligonucleótidos de doble cadena marcados con biotina.
- 4 lavados con 20 mL de 1X *Washing Buffer* de 5 minutos cada uno.
- 30 mL de *Lightshift Substrate Equilibration Buffer* durante 5 minutos.
- Por último, la membrana se incuba en 12 mL de solución de trabajo la cual contiene el sustrato de la peroxidasa y un *enhancer* de la señal luminiscente (1:1), durante 5 minutos, sin agitación y en oscuridad.

Finalmente, el revelado de las bandas se realiza mediante exposición a una película de rayos X.

**Tabla 3: Oligonucleótidos empleados para los ensayos EMSA.**

Polimorfismo	Oligonucleótidos
rs4731530	AAGGCCCCAGTTCT[T/C]GAGACAATTTACTTCCACGG
rs6950728	ACACTTTTATCATT[A/G]TAAAATGTCCCTCTTTATC
rs11982901	AGACAGTCTCACTCTG[T/C]CACCCAGGCTGGAGTGCA
rs4728141	TTTCTTTG[T/C]TGTCCAGGCT
rs13245639	ACTCCTGACCTCAGG[T/C]GATCTGTCCGCTTGGCCT
rs729302	TGCCATGGGACAAGGTGAAGAC[A/C]CAGGATCACC
rs729068	ACACGCACAGGTGTG[T/C]AGGTGCCCAGCTCA
rs12706860	GTAGCCAGGGAGCCT[G/C]CTGGACAT
rs7808659	TTCTATTTAAAAAA[A/C]ATTACTTAGGGCCGGGTA
rs754284	GCTCCCTGGCTCTGCCTG[G/C]CCCAGGAATCAGGCA
rs7801838	AAGAAATGGGGAGGAAGG[T/C]CCAGCTGTG
rs3778754	GCTTGGTAACATCATCC[C/G]CCTTTTAGAG
rs11269962	Alelo I: AAGAAGGCTCTTCTTAGCTATTGCTCCAGATACGACCAGC Alelo D: AAGAAGGCTCTTCAGATACGACCAGC
rs3757388	CTCCCAAGTAGGT[G/A]GAACTATAGG
rs3757385	AAGCGAAGAACATTCCA[T/G]GAGAAGGAACAGGAGGT
rs3807135	CTCGCCAGGGGTG[T/C]AGGCAGGCGAGAGGAGGGCC
CGGGG indel	Alelo D (3X): AGTGGATTGCGGGGCGGGGCGGGGCACTGCCCCGCGC Alelo I (4X): AGTGGATTGCGGGGCGGGGCGGGGCGGGGCACTGCC
rs3807307	AGAGAGGGCTCGGCTG[T/C]ATCGATGTAGGAACT
rs3823536	CTTACTAGGACTGCTGT[G/A]GCACTTCTCCCCTAG
rs3778753	
rs3778752	GGGCAGAAGGGAGCTTCT[A/G]GGAG[G/T][A/T]TGGATGCTGT
rs3778751	

**Tabla 3: Oligonucleótidos empleados para los ensayos EMSA.** Todas las secuencias de los oligonucleótidos de cadena sencilla están escritas en sentido 5'-3'. El marcaje se realizó con biotina en el extremo 5' de los oligonucleótidos. Entre corchetes: alelo menor/alelo mayor. I: inserción. D: delección. Los SNPs rs3778753, rs3778752 y 3778751 se analizaron en un mismo ensayo EMSA debido a su cercanía.

## 8. Elaboración de vectores recombinantes de luciferasa.

Para llevar a cabo los ensayos de gen *reporter* de luciferasa, el primer paso consistió en la elaboración de los constructos o vectores recombinantes. Se produjeron constructos para cada uno de los dos alelos de los 6 polimorfismos analizados: rs12706860, rs3778754, rs13245639, rs3807307, CGGGG indel y el indel rs11269962.

Inicialmente amplificamos los insertos (entre 150-250 pb de tamaño) a partir de individuos heterocigotos para cada uno de los polimorfismos. Los *primers* utilizados son los

mismos descritos en la **tabla 2** exceptuando para los polimorfismos: CGGGG indel (*forward*: 5'-cgccgtctggcatctccct-3'; *reverse*: 5'-tgagctctgccaggtgc-3') y el indel rs11269962 (*forward*: 5'-ctctaaggaaggggcctcac-3'; *reverse*: 5'-tggaaccttttactgga-3'). Estos insertos amplificados fueron clonados en el vector *pBluescript* (pBK SK-) mediante el protocolo de clonado TA. Este tiene como principal característica la localización de su sitio de clonación múltiple (MCS, *multicloning site*) dentro del gen *lac z* por lo que las colonias positivas se distinguen de las negativas con la hidrólisis del X-gal.

El clonado TA se basa en que la Taq polimerasa añade un residuo A al extremo 3' del producto de amplificación. Para poder utilizar esta técnica, el vector pBK SK- se linearizó previamente mediante digestión con EcoRV (New England Biolabs) y se le añadió T a sus extremos 3'. La adición de T se lleva a cabo en una reacción de 50 µL de volumen final que contiene el vector pBK SK- linearizado, 3U de *TaqDNA polymerase* (ROCHE), 2 mM de dTTP y 1X PCR buffer de reacción (ROCHE) con una concentración 1,5 mM de MgCl<sub>2</sub> y 500 mM de KCl (pH = 8,3) incubada a 72°C durante 2 horas.

La reacción de ligación tuvo lugar en 10µL de volumen final con 50 µg del vector pBK SK- linearizado y con T en sus extremos 3', 25 µg del inserto amplificado, 1 µL de *T4 DNA ligase* (New England Biolabs) y 1X *T4 DNA Ligase Reaction Buffer* con 50 mM Tris-HCl, 10 mM de MgCl<sub>2</sub>, 1 mM de ATP, 10 mM de DTT (dithiothreitol) (pH = 7,5). La reacción de ligación se incubó a 16° C toda la noche. Las bacterias *E. coli XL1-Blue* (Stratagen) fueron transformadas con los plásmidos mediante el siguiente protocolo: congelación en nitrógeno líquido durante 5 minutos, *shock* térmico en baño a 37°C durante 5 minutos, incubación en hielo durante 5 minutos y finalmente una fase de crecimiento de 1 hora en agitación a 200 rpm y 37°C después de la adición de 1 mL de medio enriquecido con glucosa 20mM. Las bacterias transformadas se crecieron en placas de agar con 100 mM de IPTG (isopropil-β-D-1-tiogalactopiranosido) y 50 mg/mL de X-gal (5-bromo-4-cloro-3-indolil-β-D-galactopiranosido) a 37°C durante toda la noche.

Gracias a la presencia del IPTG y X-gal seleccionamos las colonias de bacterias transformadas. El IPTG es un análogo no-hidrolizable de la alolactosa la cual es un metabolito isómero de la lactosa que inicia la transcripción del operón *lac*<sup>\*3</sup>. La molécula de IPTG es absorbida por la bacteria a través de la acción de la enzima permeasa de lactosa. Una vez dentro de la célula, induce la transcripción del gen *lac z* que codifica la β-galactosidasa, una hidrolasa que cataliza la hidrólisis de β-galactósidos a monosacáridos. Por otro lado, el X-gal es un análogo de galactosa que puede ser metabolizado por la β-galactosidasa en galactosa, un

<sup>\*3</sup> El operón *lac* es una unidad genética funcional requerida para el transporte y metabolismo de la lactosa en la bacteria *Escherichia coli*, así como en algunas otras bacterias entéricas. Presenta tres genes estructurales adyacentes (*lac z*, *lac y* y *lac a*), un promotor, un terminador y un operador. Es regulado por varios factores, incluyendo la disponibilidad de glucosa y de lactosa.

monosacárido, y 5-bromo-4-cloro-3-hidroxiindol, un compuesto de color azul. Las bacterias transformadas por plásmidos con el gen *lac z* intacto se colorean de azul. Sin embargo, las colonias de células transformadas con un plásmido recombinante con el gen *lac z* interrumpido por el inserto mantienen su color blanco. Se seleccionan las colonias transformadas de color blanco, se crecieron y se extrajo el ADN plasmídico mediante miniprep. El protocolo de miniprep consistió en una precipitación con NaOH seguida de una extracción con fenol-cloroformo. La clonación se comprobó mediante secuenciación.

Se subclonan los insertos mediante interés mediante doble digestión con los enzimas de restricción KpnI/SmaI (New England Biolabs) para introducirlos en un vector fos-pGL3luc. Este vector contiene el promotor mínimo<sup>\*4</sup> de *fos* localizado justo *upstream* del inicio del gen de la luciferasa de la luciérnaga (*Firefly luciferase gen*) y un gen de resistencia a la ampicilina. Tanto los insertos digeridos como el vector linearizado con los mismos enzimas se purifican mediante separación electroforética en gel de agarosa 0,8%, seguida de una purificación en columna con *Quantum Prep™ Freeze 'N Squeeze DNA Gel Extraction Spin Columns* (BIORAD).

La segunda reacción de ligación se basa en los extremos cohesivos dejados por los enzimas de restricción. Esta segunda reacción y la de transformación se realizaron siguiendo los protocolos descritos previamente. La selección de bacterias recombinantes se basó en la resistencia a la ampicilina.

## 9. Transfección: electroporación.

Para llevar a cabo los ensayos gen *reporter* de luciferasa aplicamos la metodología de electroporación como método de transfección. La electroporación funciona por la formación de poros en la membrana plasmática de la célula, que se sellan tras un corto período de tiempo, durante el cual los plásmidos recombinantes tienen la oportunidad de entrar a la célula. La electroporación se lleva a cabo en un electroporador, un aparato que crea un campo electromagnético regulado a través de la suspensión celular.

En nuestro caso transfectamos las células WIL2 NS mediante el electroporador *Microporator MP-100* y el kit *Neon® Transfection System* (Invitrogen). El protocolo consistió

<sup>\*4</sup> El promotor mínimo de eucariotas se define como la secuencia de ADN mínima requerida para que se inicie una transcripción correcta *in vivo*. Típicamente, consta de 40 pb localizadas *upstream* del punto de inicio de la transcripción. Sus componentes son cuatro, ordenados en dirección 5' a 3': (1) TFIIB *recognition element*, secuencia reconocida por el factor de inicio de la transcripción TFIIB. (2) Caja TATA, elemento reconocido por la proteína de unión a la caja TATA o TBP (TATA *binding protein*). (3) Inr o iniciador, que alberga al punto de inicio de la transcripción y al cual se une el factor TFIID, y, probablemente, la propia ARN polimerasa II. (4) DPE (*downstream promoter element*), reconocido por TFIID.

en la transfección de  $2 \times 10^6$  células con 4  $\mu\text{g}$  de ADN. Los vectores utilizados en nuestros experimentos fueron los siguientes: dos plásmidos recombinantes del vector fos-pGL3luc con los alelos para cada polimorfismo a estudiar, el vector fos-pGL3luc sin inserto, el vector pRL-TK Renilla (Promega) que expresa el gen que codifica para la luciferasa renilla-luciferina 2-monooxigenasa que proviene de *Renilla reniformis*, pBK SK- y el vector RSV-luc que proviene del adenovirus E1 y expresa el gen de la luciferasa bajo el control del promotor CMV\*<sup>5</sup>. Nuestro experimento consta de 5 puntos: blanco, control positivo de transfección (RSV-luc: 400 ng; pBK SK-: 2800 ng y pRL-TK: 800 ng), vector vacío (fos-pGL3luc vector: 3200 ng; pRL-TK: 800 ng), alelo 1 (alelo1.fos-pGL3luc vector: 3200 ng; pRL-TK: 800 ng) y alelo 2 (alelo2.fos-pGL3luc vector: 3200ng; pRL-TK: 800 ng).

Las células se lavan con PBS. Se resuspenden en 100  $\mu\text{L}$  de *Resuspension buffer R* junto con los 4  $\mu\text{g}$  de ADN especificados por punto. Seguidamente, la electroporación se lleva a cabo con: un pulso de 1100 voltios durante 30 mili segundos. Después, las células se siembran en placas de 6 pocillos con 2 mL de medio completo RPMI y se cultivan a 37 °C en incubador con 5% CO<sub>2</sub> durante 24 horas. A partir de aquí realizamos el experimento con células no estimuladas y estimuladas con PMA e ionomicina. Las condiciones de estimulación consisten en suplementar al medio de cultivo con 500 ng/mL de *ionomycin* y 10 ng/mL de PMA (*phorbol 12-myristate 13-acetate*) durante 4 horas. Por último, para el revelado de los experimentos, utilizamos los kits *Firefly Luciferase Assay System* y el *Renilla Luciferase Assay System* (Promega) siguiendo el protocolo especificado por el fabricante.

## 10. Análisis estadístico:

Para la realización de los análisis estadísticos se utilizaron los siguientes programas: Statistica 7.0 (StatSoft, Tulsa, OK, USA), Haploview<sup>1</sup>, Phase 2.1<sup>2</sup> y R. Para el cálculo y representaciones gráficas de los valores de desequilibrio de ligamiento entre polimorfismos ( $r^2$  y  $D'$ ) se utilizó el programa Haploview. La concordancia con el equilibrio Hardy-Weinberg se determinó en las muestras control con el test  $\chi^2$  *goodness-of-fit* considerando el umbral de significación de 0,05 si corregir por múltiples tests.

\*<sup>5</sup> Este promotor es el de la región inmediata temprana de Citomegalovirus (CMV/IE), el cual permite altos niveles de transcripción de manera constitutiva en una gran variedad de células eucariotas.

En todos los análisis se consideró un modelo aditivo con valores 0, 1 y 2 para los genotipos AA, Aa y aa respectivamente. Por otro lado, la estimación de haplotipos se llevó a cabo mediante el programa Phase 2.1 (con los parámetros que el programa usa por defecto) que utiliza un algoritmo Bayesiano.

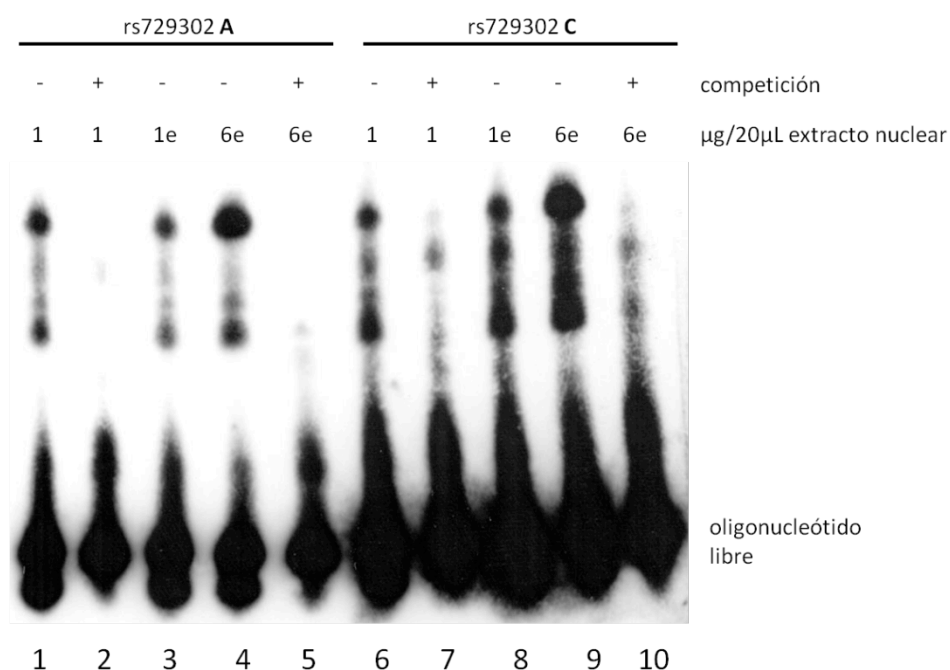
En relación a los ensayos de cambio de movilidad electroforética (EMSA, *Electroforetic Mobility Shift Assay*) incluidos en este apartado, solamente se consideraron resultados válidos y reproducibles aquellos que se replicaron en tres experimentos independientes. Por otro lado, para el análisis de los resultados de los ensayos *gen reporter*, lo primero que hicimos fue referenciar las unidades de luminiscencia de *Firefly* luciferasa por la actividad de la *Renilla* luciferasa para así ajustar todos los puntos del experimento en base a la eficiencia de la transfección. Esta normalización se llevó a cabo dividiendo los datos de luminiscencia de *Firefly* luciferasa entre los datos de luminiscencia de *Renilla* luciferasa por punto y finalmente multiplicando este resultado por 100. Para comparar los niveles de expresión de cada uno de los alelos por polimorfismo estudiado se utilizó el test de Wilcoxon de datos pareados. Para ello, los datos analizados son las unidades luminiscencia de *Firefly* luciferasa normalizadas por cada alelo referenciadas al vector vacío (*fos-pGL3-Basic*). Realizamos 5 experimentos independientes por polimorfismo analizado de los cuales los resultados para cada par de alelos se consideran medidas pareadas.

En relación a todos los test estadísticos empleados, se consideró el umbral de significación de 0,05 sin corregir por múltiples test.



## 1. Ausencia de diferencias en EMSA entre los dos alelos del SNP rs729302.

A partir de los resultados obtenidos en nuestro estudio previo (Publicación 3), decidimos analizar el efecto protector del gen *IRF5* frente al LES asociados con el SNP rs729302. Este polimorfismo podría influir en la regulación de su expresión génica ya que se encuentra en la región promotora, 9 Kb *upstream* del inicio del gen. Para testar esta posibilidad, realizamos un ensayo EMSA con extractos nucleares procedentes de la línea celular WIL2 NS tanto sin estimular como estimulados con ionomicina y PMA (Tabla 3 de *Material y Métodos*). Estos tests no mostraron diferencias entre los dos alelos (Figura 1).



**Figura 1: Ensayo EMSA con los dos alelos del SNP rs729302.** Ensayo EMSA con 1 µg y 6 µg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

## 2. Mapa de desequilibrio de ligamiento (LD).

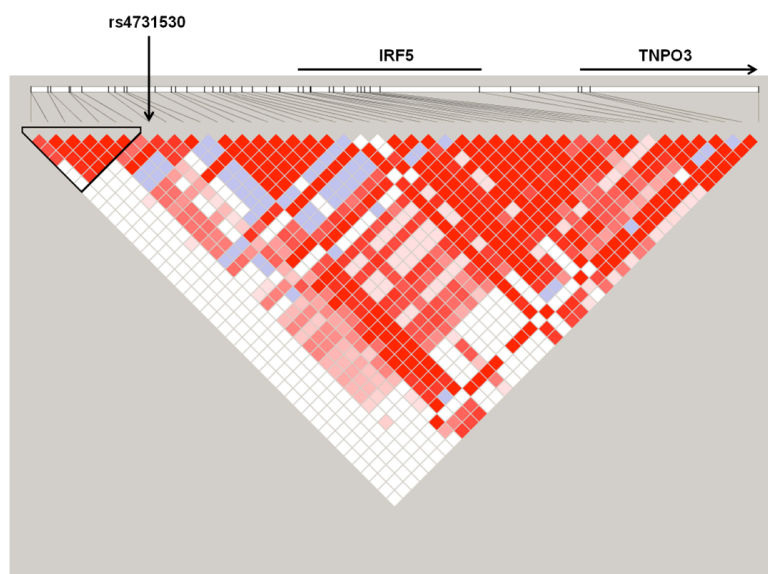
Debido a la falta de diferencias en el EMSA con los dos alelos del SNP rs729302, lo más probable es que este SNP funcione como marcador de otro polimorfismo. Por lo tanto,

decidimos estudiar la región promotora del gen *IRF5*. Para ello, realizamos un mapa de LD que abarca el gen *IRF5* en su totalidad, 21,5 Kb de su región promotora (referenciado al inicio del exón 1 variante A) y parte del gen *TNPO3* hasta unas 50 Kb. Para ello genotipamos 54 polimorfismos (**Tabla 1 de Material y Métodos**): los 11 polimorfismos analizados en nuestro estudio previo (**Publicación 3**) y 43 en *dbSNP build 128* con un status *doublehit* de validación. Se genotiparon en 95 muestras de controles españoles.

En los análisis se excluyeron los SNPs con una frecuencia de su alelo menor  $< 2\%$  (8 SNPs: rs12536195, rs6951615, rs11773414, rs11983607, rs11763323, rs754280, rs41298401 y rs11767834) y los SNPs rs3778752 y rs3778751 ya que estuvieron en completo LD ( $r^2 = 1$ ) con el SNP rs3778754. Por lo tanto, el número total de polimorfismos en las gráficas y tablas es 44.

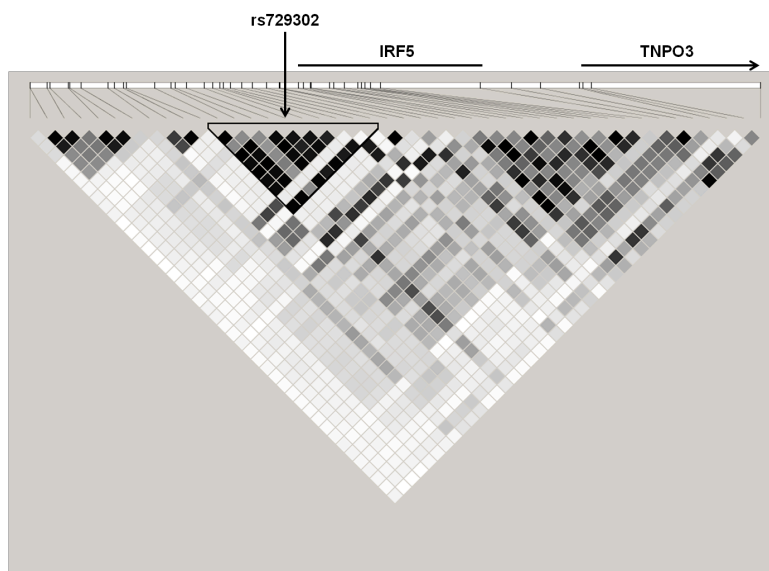
Para la elaboración del mapa de LD se utilizó el programa Haploview<sup>300</sup>. En la **tabla suplementaria 1** se representan los valores de  $D'$  y  $r^2$  para todas las combinaciones dos a dos de los polimorfismos estudiados.

El mapa de  $D'$  muestra que en la región *upstream* al rs4731530 había un bloque con alta  $D'$  e independiente de la región que comprendía al gen *IRF5* y su promotor más proximal. Por lo tanto, los 7 SNPs de este bloque (rs6955705, rs4731528, rs10081379, rs10280295, rs6970780, rs6951243 y rs960633) fueron excluidos del estudio de la distribución haplotípica. Por último, se debe mencionar que el bloque de LD de *IRF5* se extendía hasta el gen de la *TNPO3* (**Figura 2**).



**Figura 2: Mapa de LD representado en valores de  $D'$ .** Se incluyeron los 44 polimorfismos seleccionados. La escala de blanco a rojo indica valores de  $D'$  de 0 hasta 1.

El mapa de LD con los valores de  $r^2$  mostró, una zona de alto LD con el rs729302 que está compuesta por 8 SNPs con un  $r^2 > 0,7$  con el rs729302: rs6950728, rs11982901, rs13245639, rs729068, rs12706860, rs7808659, rs754284 y rs7801838 (**Figura 3**).



**Figura 3: Mapa de LD representado en valores de  $r^2$ .** Se incluyeron los 44 polimorfismos seleccionados. La escala de blanco a negro indica valores de  $r^2$  de 0 hasta 1.

### 3. Búsqueda de polimorfismos funcionales en la región promotora de *IRF5*.

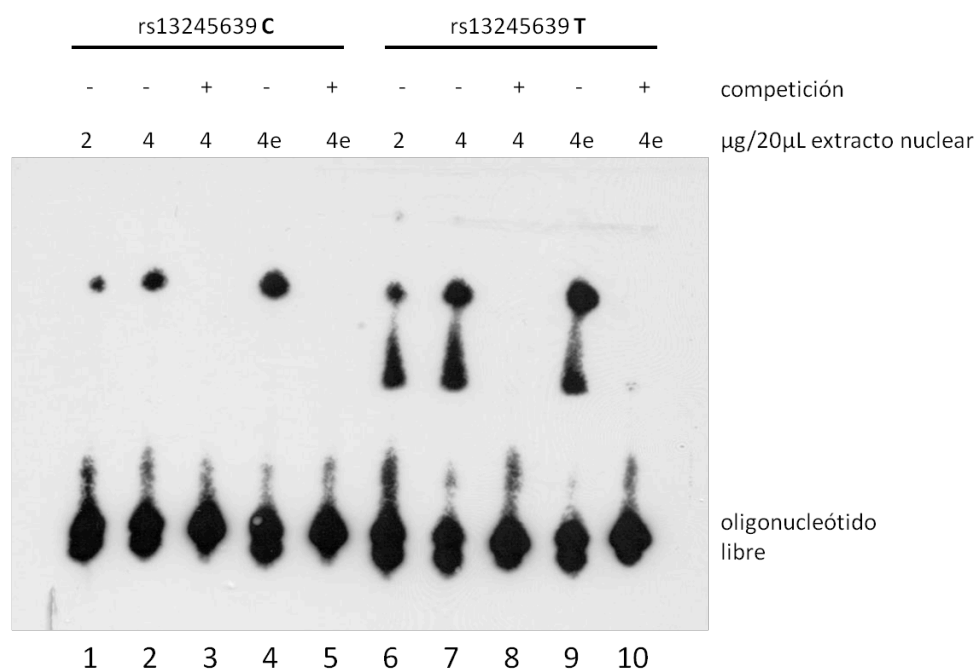
#### 3.1. Bloque de alta correlación con el SNP rs729302:

Se seleccionaron los 8 SNPs con correlación elevada con el rs729302 para analizar mediante ensayos EMSA la unión de sus alelos con los extractos nucleares. (**Tabla 1**).

**Tabla 1: SNPs en alto LD con el SNP rs729302.**

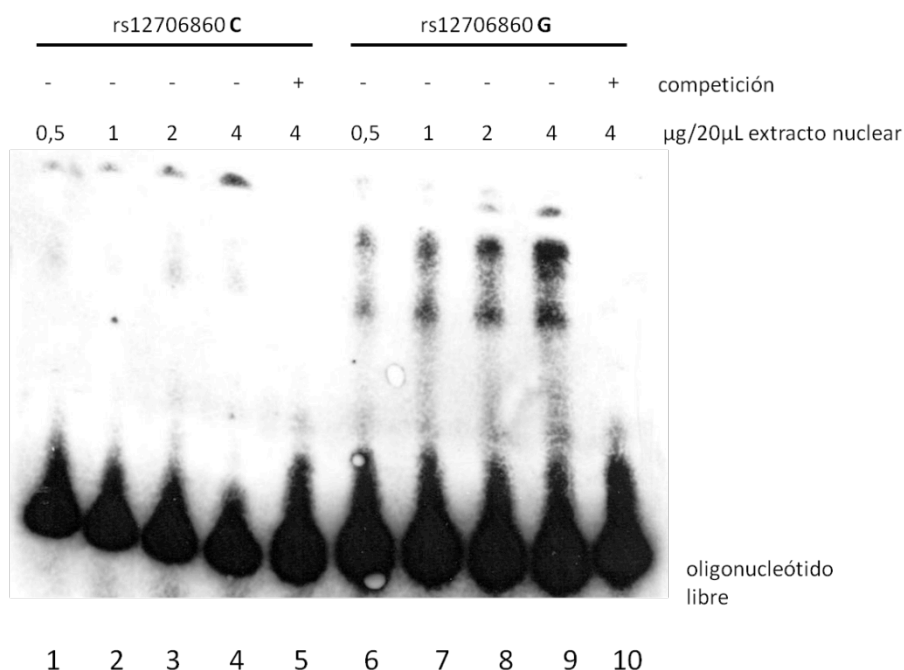
SNP	Posición	Alelos	$r^2$ con rs729302
rs6950728	128352897	A < G	0,97
rs11982901	128353990	T < C	0,97
rs13245639	128355046	T < C	0,97
rs729068	128356797	T < C	0,90
rs12706860	128357262	G < C	0,83
rs7808659	128357474	A < C	0,90
rs754284	128358714	G < C	0,87
rs7801838	128361230	T < C	0,73

De los 8 SNPs solamente dos mostraron unión alélica diferencial: rs13245639 y el rs12706860. Los ensayos EMSA de los 6 SNPs restantes se muestran en las **figuras suplementarias 1-6**.



**Figura 4: Ensayo EMSA de los alelos de rs13245639.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

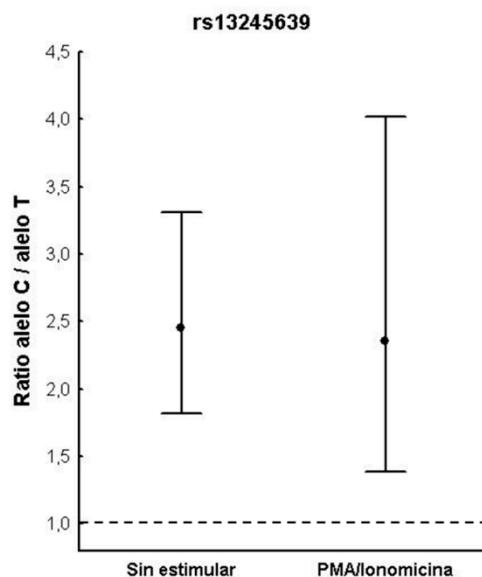
En el ensayo EMSA del SNP rs13245639 observamos diferencias en la unión de proteínas nucleares. Concretamente, el alelo menor T mostró una banda adicional en el patrón de unión que estaba ausente cuando utilizamos los oligonucleótidos con el alelo C (**Figura 4**). Esta unión era específica ya que desaparecía en los ensayos de competición correspondientes.



**Figura 5: Ensayo EMSA de los alelos de rs12706860.** Ensayo EMSA con extracto nuclear de la línea celular WIL2 NS sin estimular con concentraciones de 0,5μg a 4μg. En los ensayos de competición (carriles 5 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

También se observaron diferencias en la unión de proteínas nucleares con los alelos del SNP rs12706860. En el caso del alelo C se encontró una banda de unión específica en la zona superior de los geles que era de menor de intensidad con el alelo G. También se encontró un patrón de bandas de unión específico para el alelo menor G que no apareció con el alelo C (**Figura 5**). Resultados similares se obtuvieron con extractos nucleares procedentes de la línea celular WIL2 NS estimulada (datos no mostrados).

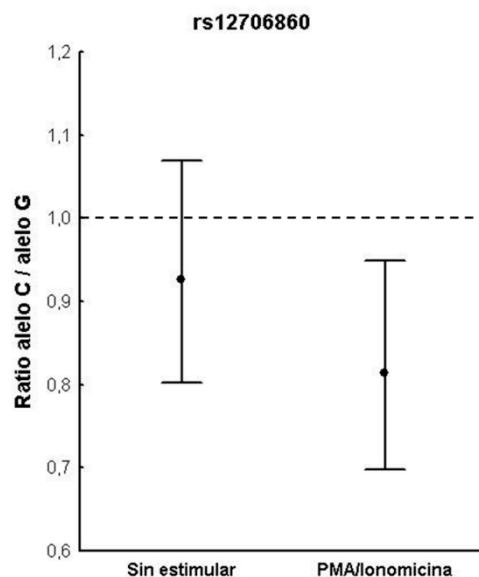
Estos resultados pueden indicar que los SNPs rs13245639 y rs12706860 podrían presentar cambios en la unión de factores de transcripción que afecten a la regulación de la expresión de *IRF5*. Para testar esta hipótesis realizamos ensayos gen *reporter* de luciferasa con constructos para cada uno de los alelos de los dos SNPs.



**Figura 6: Ensayo de gen *reporter* para los dos alelos del SNP rs13245639.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo C frente al alelo T. Se realizaron 5 experimentos independientes.

Los insertos con cada uno de los alelos se insertan *upstream* del promotor en el vector *fos-pGL3luc*. Solamente el SNP rs13245639 mostró diferencias significativas entre sus dos alelos. El alelo C mostró una mayor expresión del *reporter* que el alelo T (**Figura 6**). Este resultado se observó tanto en condiciones basales como tras estimulación con PMA e ionomicina (Wilcoxon test  $P = 0,043$  con ambas condiciones). Teniendo en cuenta los resultados del ensayo EMSA (**Figura 4**) y ensayo gen *reporter* de luciferasa (**Figura 6**), parece que la unión específica que se observó con el alelo T en el EMSA es con un represor de la transcripción.

Los ensayos *reporter* del rs12706860 mostraron una tendencia hacia una disminución de la expresión promovida por el alelo C en relación al alelo G bajo condiciones de estimulación con PMA e ionomicina (**Figura 7**). Sin embargo, ninguna de las dos comparaciones, tanto en condiciones basales como tras estimulación, fueron significativas (Wilcoxon test  $P = 0,5$  en condiciones basales y  $P = 0,08$  bajo estimulación). Por lo tanto, no parece que la unión alélica diferencial encontrada en el EMSA previo (**Figura 5**) de lugar a un cambio significativo en la regulación de la expresión.



**Figura 7: Ensayo de gen *reporter* para los dos alelos del rs12706860.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo C frente al alelo G. Se realizaron 5 experimentos independientes.

### **3.2. Estudio de los haplotipos #1 y #2 (protectores):**

Como los haplotipos protectores encontrados en nuestro estudio de asociación (#1 y #2) pueden identificar otros polimorfismos causales, realizamos una estimación de los haplotipos con los 37 polimorfismos con frecuencia del alelo raro > 2% y tras excluir los 7 más distales 5' al gen (que mostraban independiente LD en el análisis de D'). Este análisis mostró 20 haplotipos de frecuencia mayor al 1% (**Tabla 2**) que incluyen el 85,3% de los cromosomas analizados.

Algunos de los 6 haplotipos de nuestro estudio de asociación previo (**Publicación 3**), se subdividieron en subhaplotipos de menor frecuencia debido a la inclusión de más polimorfismos. Este es el caso del haplotipo de protección #2, que se subdividió en tres subhaplotipos de frecuencias 7,9%, 3,2%, 1,1%. Los haplotipos #1 y #5 se mantuvieron. Se encontraron 7 nuevos haplotipos con frecuencias entre el 1,1% al 3,2 % marcados como “nuevo” en la **tabla 2**. A partir de esta distribución de haplotipos, se seleccionaron 9 SNPs de la región promotora de *IRF5* cuyos alelos raros agruparon la mayoría de los haplotipos y subhaplotipos #1 y #2, aunque también incluyeron neutros o nuevos (**Tabla 2**).

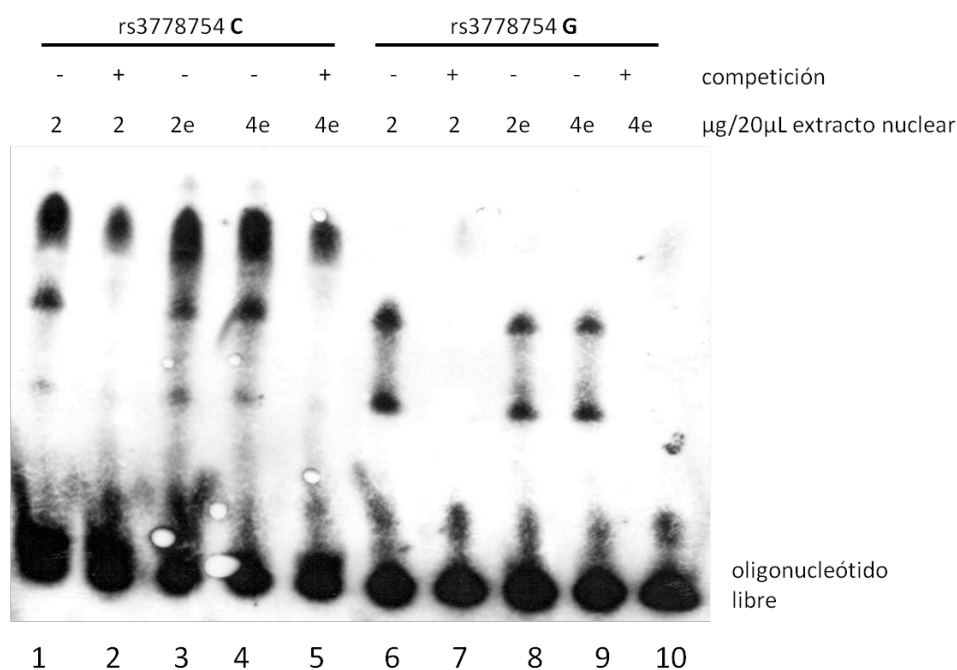
Hap.	%	rs4731530	rs2402941	rs6968225	rs6968508	rs6950728	rs11982901	rs4728141	rs13245639	rs729302	rs729068	rs12706860	rs7808659	rs2402940	rs754284	rs754281	rs11768806	rs4728142	rs7801838	rs1874330	rs3778754	rs3757388	rs3757387	rs3757385	rs3807135	rs2004640	rs3807307	rs752637	rs3823536	rs3778753	rs3807306	Indel exón 6	rs10954213	rs13242262	rs10488630	rs10488631	rs2280714	rs4731535
#4	24,7	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	C	A	G	T	D	A	T	G	T	T	T
#2	7,9	T	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	G	T	T	C	G	T	T	T	G	T	T	G	A	G	I	G	A	A	T	C	C
#3	7,9	T	A	G	G	G	C	T	C	A	C	C	C	T	C	T	T	G	C	C	C	G	T	T	T	G	T	T	G	A	G	I	G	A	A	T	C	C
#6	6,8	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	C	A	G	T	I	A	T	A	C	T	C
#1	5,3	T	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	G	T	T	C	A	T	G	C	G	T	C	C	G	A	G	D	A	T	A	T	T
#4	3,2	T	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	C	A	G	T	D	A	T	G	T	T	T
nuevo <sup>a</sup>	3,2	T	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	G	T	T	C	G	T	T	T	T	G	T	T	G	A	T	D	A	T	G	T	T
#5	3,2	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	C	A	G	T	D	A	T	A	T	T	T
nuevo <sup>a</sup>	3,2	C	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	G	T	T	C	A	T	G	C	G	T	C	C	G	A	G	D	A	T	G	T	T
#2	3,2	C	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	G	T	T	C	G	T	T	T	G	T	C	C	G	A	G	I	G	A	A	T	C
#3	2,6	T	G	C	C	G	C	C	C	A	C	C	C	T	C	T	T	G	C	C	C	G	T	T	T	G	T	T	T	G	A	G	I	G	A	A	T	C
nuevo	2,6	C	G	C	C	G	C	T	C	A	C	C	C	C	C	C	C	G	C	T	C	A	T	G	C	T	T	C	C	G	A	G	I	G	A	A	T	C
#6	2,6	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	G	C	T	G	A	C	G	C	T	C	C	A	A	G	T	I	A	T	A	C	T
nuevo <sup>a</sup>	1,6	T	G	C	C	A	T	T	T	C	T	G	A	C	G	C	C	A	C	T	G	A	C	G	C	T	C	C	A	A	G	T	D	A	T	G	T	T
nuevo	1,6	C	G	C	C	G	C	T	C	A	C	G	C	C	C	C	C	G	C	T	G	A	T	G	C	T	C	T	A	A	G	T	D	A	T	A	T	T
nuevo	1,6	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	T	A	A	G	T	D	A	T	G	T	T
#2	1,1	T	G	C	C	A	T	T	T	C	T	G	C	C	G	C	C	G	T	T	C	G	T	T	T	G	T	T	G	A	G	I	G	A	A	T	C	C
#3	1,1	T	A	G	G	G	C	T	C	A	C	C	C	T	C	C	C	G	T	T	C	G	T	T	T	G	T	T	G	A	G	I	G	A	A	T	C	C
nuevo	1,1	C	G	C	C	G	C	T	C	A	C	C	C	C	C	C	C	G	C	T	C	A	T	G	C	T	T	T	G	A	G	I	G	A	A	T	T	C
#4	1,1	C	G	C	C	G	C	C	C	A	C	C	C	C	C	C	C	A	C	T	G	A	C	G	C	T	C	C	A	A	G	G	D	A	T	G	T	T

**Tabla 2: Distribución haplotípica de los 37 polimorfismos analizados.** Los SNPs analizados en nuestro estudio de asociación previo (**Publicación 3**) están resaltados en negrita. La numeración de los haplotipos se mantiene. nuevo: haplotipos nuevos. nueva: haplotipo nuevo con el alelo menor C del SNP protector rs729302. Los SNPs seleccionados debido a su alta  $r^2$  con el SNP protector rs729302 están subrayados. Los SNPs seleccionados están marcados en gris.



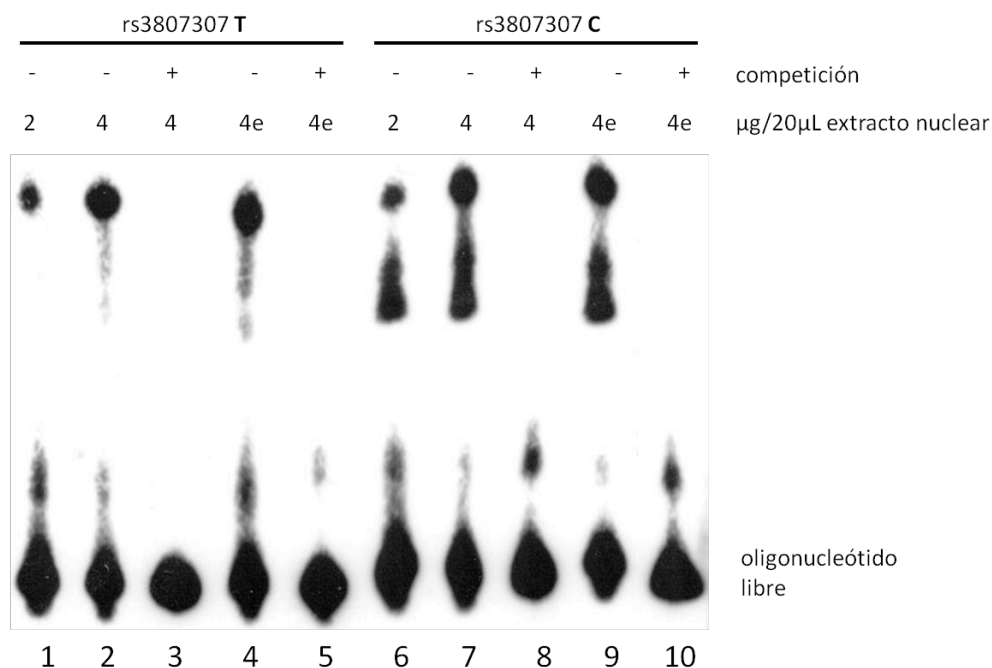
Los 9 SNPs seleccionados a partir de los haplotipos fueron: rs4731530, rs4728141, rs3778754, rs3757388, rs3757385, rs3807135, rs3807307, rs3823536 y el rs3778753. Todos ellos fueron analizados mediante EMSA para determinar la unión diferencial de sus alelos con extractos nucleares. El SNP rs3778754 estaba en completo LD con dos SNPs rs3778752 y rs3778751. Estos dos SNPs fueron también analizados en EMSA al incluirlos en los mismos oligonucleótidos con el rs3778753 al que están muy próximos. Solamente, dos de los SNPs mostraron unión alélica diferencial en los EMSA: rs3778754 y rs3807307. Los EMSA de los 7+2 SNPs restantes no mostraron diferencias y se muestran en las **figuras suplementarias 7-13**.

El ensayo EMSA de los dos alelos de rs3778754 podemos mostraba una banda de unión adicional con el alelo menor C que está totalmente ausente con el alelo G (**Figura 8**). Este fenómeno se observa tanto con extractos nucleares procedentes de células en condiciones basales como estimuladas con PMA e ionomicina.



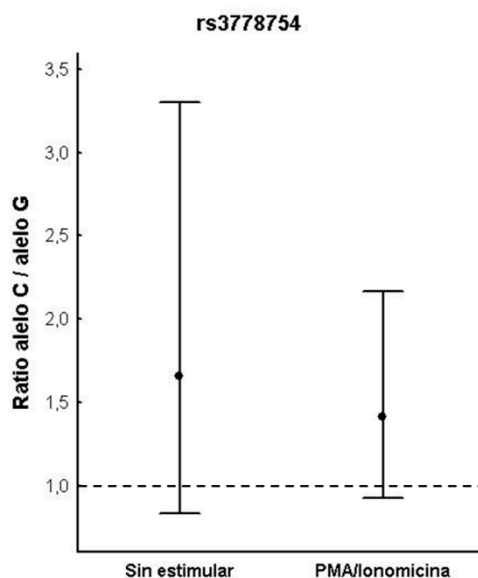
**Figura 8: Ensayo EMSA de los dos alelos de rs3778754.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

El SNP 3807307 también mostró unión alélica diferencial en el EMSA caracterizada por la aparición de una banda de unión adicional con el alelo mayor C en comparación con el alelo T (**Figura 9**). La unión es específica ya que desaparece en los ensayos de competición correspondientes. Por otro lado, no observamos ningún cambio en el patrón de unión cuando usamos extractos nucleares de las células sin estimular o estimuladas con PMA e ionomicina.



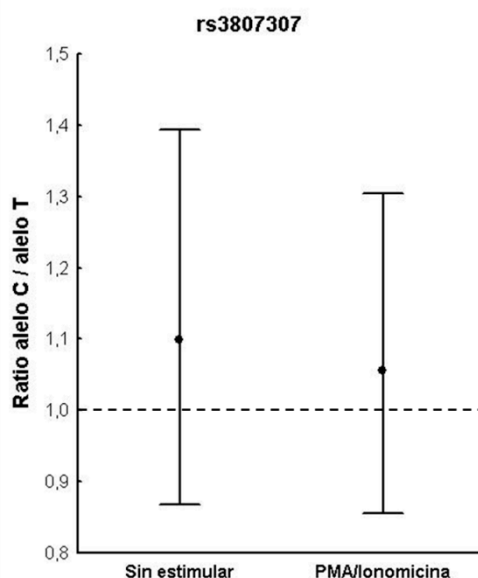
**Figura 9: Ensayo EMSA de los alelos de rs3807307.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

Seguidamente se procedió a realizar ensayos gen de *reporter* de luciferasa para analizar la repercusión en la regulación de la expresión de la unión diferencial de los SNP rs3778754 y rs3807307. Ninguno de los dos polimorfismos mostró diferencias significativas entre sus alelos.



**Figura 10: Ensayo de gen *reporter* de luciferasa para los dos alelos del SNP rs3778754.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo C frente al alelo G. Se realizaron 5 experimentos independientes.

El SNP rs3778754 no mostró ninguna diferencia significativa entre los ensayos de gen *reporter* con sus dos alelos tanto en condiciones basales como tras estimulación con PMA e ionomicina (Wilcoxon test  $P = 0,08$  en ambas condiciones) (**Figura 10**). Del mismo modo, en los ensayos de gen *reporter* de rs3807307, tampoco se observaron diferencias significativas entre los dos alelos en las dos condiciones analizadas (Wilcoxon test  $P = 0,3$  en condiciones basales y  $P = 0,5$  bajo estimulación) (**Figura 11**).



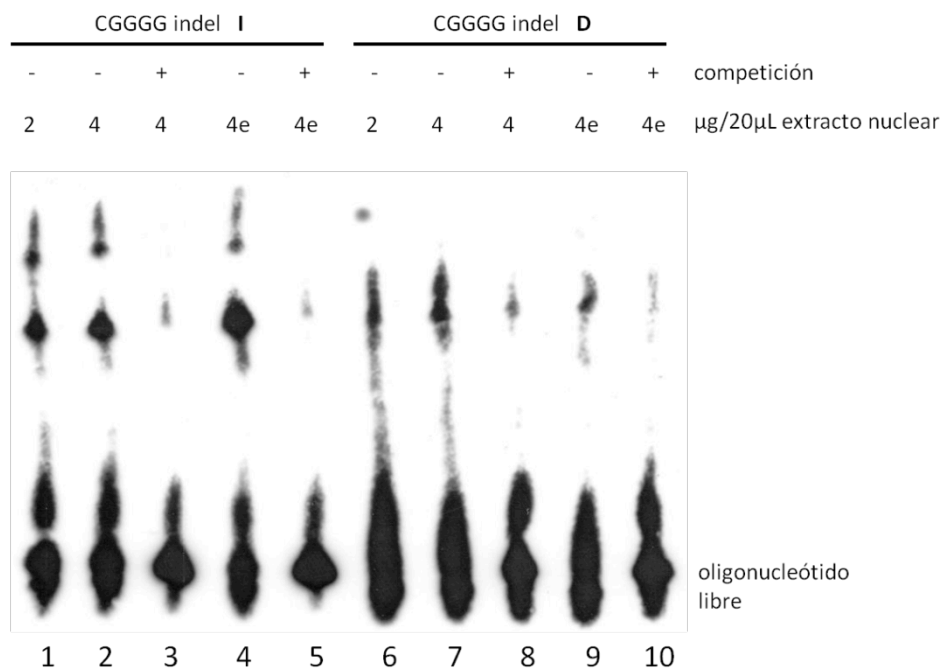
**Figura 11: Ensayo de gen *reporter* para los dos alelos de rs3807307.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo C frente al alelo T. Se realizaron 5 experimentos independientes.

#### 4. Indel CGGGG.

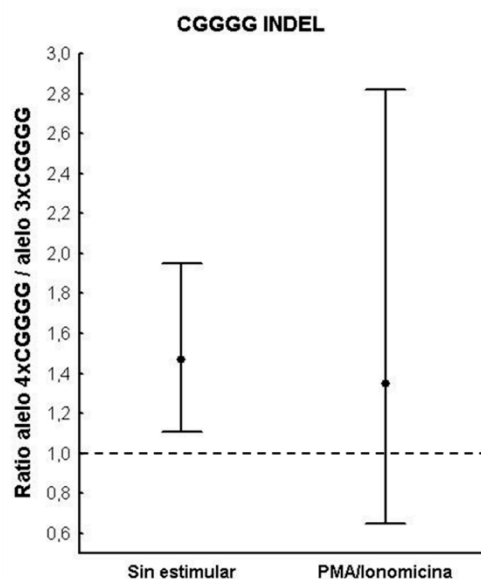
Como ya se ha señalado en la introducción, existe un polimorfismo denominado indel CGGGG con repercusión funcional conocida en la región promotora de *IRF5*. Este polimorfismo consiste en una secuencia de 5 pb (CGGGG) que se repite 4 veces (inserción) o 3 veces (delección)<sup>273, 282, 283</sup>. La inserción 4xCGGGG crea un lugar de unión adicional para el factor de transcripción SP1 (*specificity protein 1*)<sup>282, 283</sup> que determina niveles de expresión de *IRF5* entre 2 y 10 veces superiores a las células homocigotas para la delección 3xCGGGG<sup>283</sup>.

Decidimos utilizar este polimorfismo como control positivo de nuestros experimentos. Nuestros resultados con EMSA son similares a los descritos (**Figura 12**). También el resultado del experimento de gen *reporter* fue similar a los descritos (**Figura 13**) aunque no hubo

diferencias entre los alelos tras estimulación. Hay que tener en cuenta que todos los resultados publicados en relación con este polimorfismo se habían realizado con células en condiciones basales. Por lo tanto, estos resultados validan nuestro sistema de evaluación en relación con los utilizados por otros autores.



**Figura 12: Ensayo EMSA de los dos alelos del indel CGGGG.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

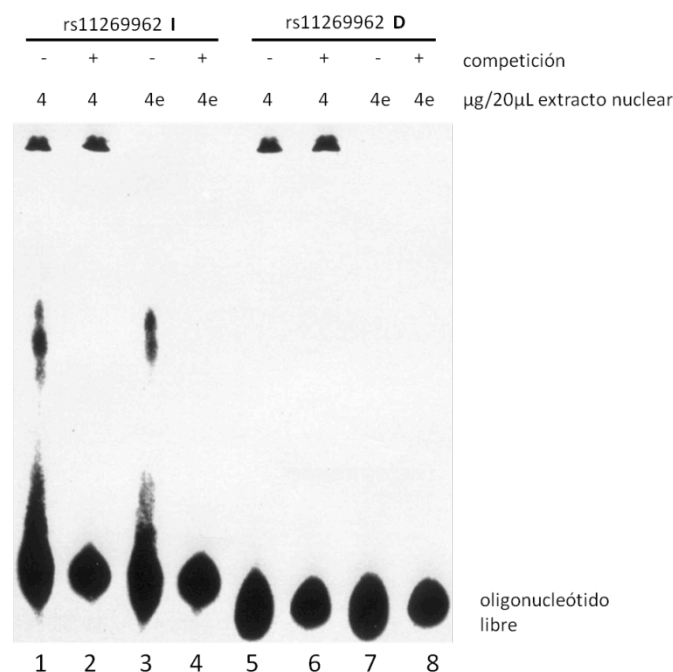


**Figura 13: Ensayo de gen *reporter* de los dos alelos del indel CGGGG.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo inserción 4xCGGGG frente al alelo delección 3xCGGGG. Se realizaron 5 experimentos independientes.

## 5. Indel rs11269962.

Un estudio de RNA-seq no publicado realizado en colaboración con la Universidad de McGill de Montreal, había identificado un polimorfismo en la región 5' a *IRF5* que se correlacionaba con la expresión del gen. Este polimorfismo, rs11269962, consiste en un indel de 14 pb y su alelo más corto o delección se asocia con un aumento de la expresión de *IRF5*. Por lo tanto, se seleccionó este polimorfismo para análisis funcionales.

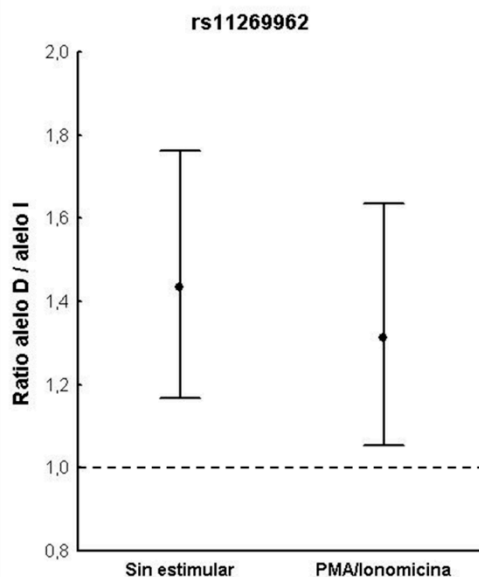
El ensayo EMSA de los dos alelos del indel rs11269962 mostró unión alélica diferencial que consiste en una banda de unión presente con el alelo inserción y ausente en el alelo delección (**Figura 14**). Esta unión adicional fue como específica ya que desaparecía completamente en los ensayos de competición correspondientes. Se observó esta misma unión alélica diferencial tanto en condiciones basales como tras estimulación con PMA e ionomicina.



**Figura 14: Ensayo EMSA de los dos alelos del indel rs11269962.** Ensayo EMSA con 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 4, 6 y 8) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

Se puede señalar que los extractos nucleares de células sin estimular presentaron una unión adicional en comparación con los extractos de células estimuladas con PMA e ionomicina. De todos modos, esta unión no fue específica ya que también aparecía en los ensayos de competición con el oligonucleótido no marcado (carriles 2 para el alelo inserción y carril 6 para el alelo delección).

El ensayo de gen *reporter* con constructos para los dos alelos del indel rs11269962 mostró que el alelo delección promueve un aumento de la expresión del gen de la luciferasa comparado con el alelo delección (**Figura 15**). Este resultado sólo fue significativo en condiciones basales (Wilcoxon test  $P = 0,04$ ). Sin embargo, los resultados obtenidos tras estimulación siguieron la misma tendencia hacia un aumento de la expresión de luciferasa con el alelo delección (Wilcoxon test  $P = 0,08$ ). Por lo tanto parece que el alelo inserción del indel rs11269962 une a un represor de la expresión.



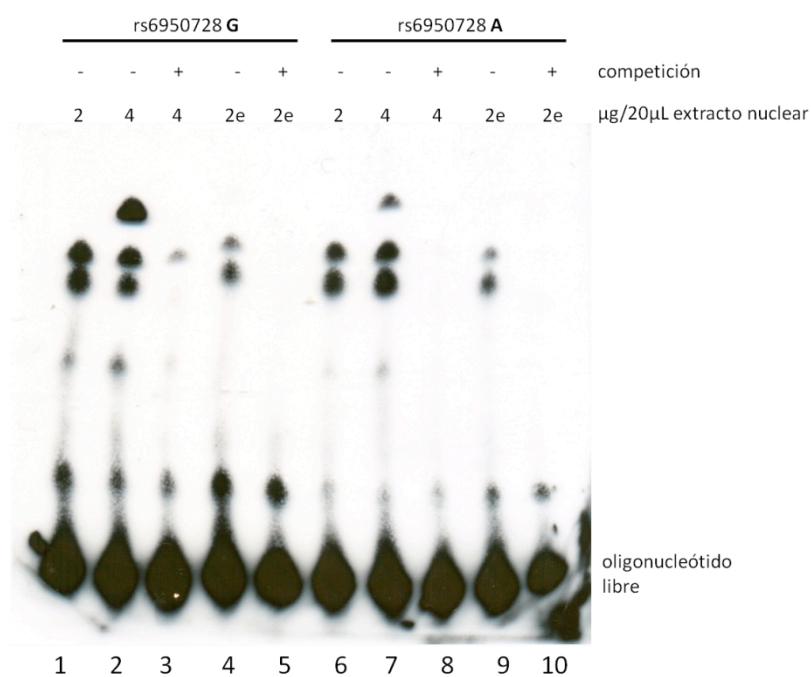
**Figura 15: Ensayo de gen *reporter* para los dos alelos del indel rs11269962.** Los dos alelos se compararon mediante una transfección transitoria en células WIL2 NS en condiciones basales y tras estimulación con PMA e ionomicina. Se representaron las medias geométricas e intervalos de confianza al 95% (95% I.C.) de los *ratios* de las señales de luminiscencia del alelo delección (D) frente al alelo inserción (I). Se realizaron 5 experimentos independientes.

En resumen, la búsqueda de polimorfismos funcionales que puedan explicar el efecto protector de la variabilidad genética en *IRF5*, resultó en la identificación de dos polimorfismos: rs13245639 y el indel rs11269962 (localizados en la región 5' al gen).

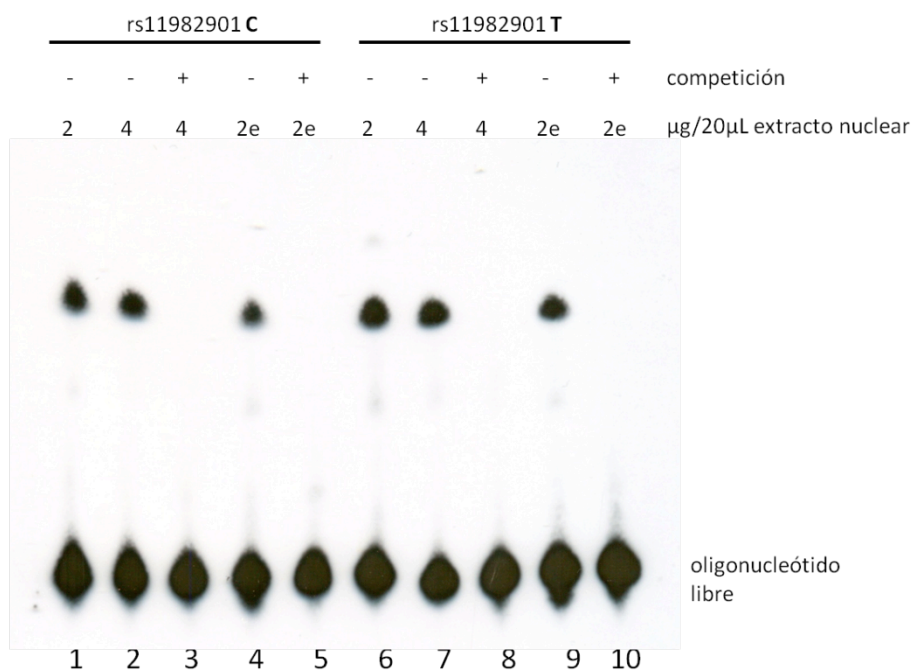
		D'																																																	
		r6955705	r6731528	r60081379	r60280295	r69707780	r6051243	r6040633	r6731530	r6402941	r69608225	r6968508	r6950728	r1992901	r6728141	r613245639	r6729302	r6729068	r67206860	r67808659	r6402940	r6754284	r6754281	r61706806	r6728142	r67801838	r61874330	r6779754	r6757388	r6757387	r6757385	r6807135	r62006640	r6807307	r6752637	r6823536	r6779753	r6807306	Indel exon 6	r6054213	r61324262	r61488630	r60488631	r6280714	r6731535						
r6955705		0.839	0.834	0.839	0.481	0.481	0.555	0.303	0.245	0.268	0.245	0.0	0.173	0.0	0.0070	0.0	0.031	0.097	0.216	0.042	0.266	0.266	0.278	0.056	0.209	0.129	0.066	0.312	0.066	0.059	0.125	0.233	0.173	0.233	0.233	0.233	0.158	0.381	0.322	0.226	0.535	0.222	0.14	0.463							
r6731528		1.0	0.934	0.928	0.928	0.888	0.24	0.293	0.165	0.236	0.236	0.359	0.359	0.22	0.359	0.371	0.359	0.284	0.297	0.066	0.331	0.107	0.107	0.041	0.293	0.055	0.215	0.373	0.171	0.373	0.072	0.125	0.223	0.197	0.217	0.155	0.06	0.22	0.214	0.535	0.482	0.314	0.0010								
r60081379		0.13	0.938		0.978	1.0	1.0	0.962	0.279	0.284	0.145	0.224	0.396	0.377	0.255	0.377	0.389	0.401	0.31	0.324	0.052	0.35	0.147	0.147	0.088	0.341	0.046	0.239	0.377	0.224	0.377	0.38	0.245	0.224	0.264	0.224	0.224	0.192	0.071	0.246	0.239	0.117	0.487	0.338	0.013						
r60280295		0.137	0.853	0.917		1.0	1.0	0.963	0.331	0.306	0.0	0.085	0.441	0.419	0.264	0.419	0.433	0.445	0.385	0.372	0.089	0.391	0.257	0.257	0.1	0.392	0.076	0.23	0.326	0.23	0.326	0.329	0.245	0.215	0.247	0.215	0.186	0.035	0.196	0.225	0.082	0.504	0.328	0.0030							
r69707780		0.024	0.455	0.503	0.525	0.953		0.976	0.95	0.771	1.0	1.0	1.0	0.752	0.752	0.676	0.752	0.742	0.752	0.706	0.724	1.0	0.761	0.487	0.487	0.391	0.717	0.62	0.58	0.702	0.555	0.702	0.703	0.593	0.579	0.626	0.579	0.579	0.53	0.308	0.538	0.532	0.18	0.354	0.258						
r6951243		0.024	0.455	0.503	0.525	0.953		0.975	0.766	1.0	1.0	1.0	0.752	0.752	0.676	0.752	0.742	0.752	0.706	0.724	1.0	0.761	0.487	0.487	0.391	0.717	0.62	0.58	0.691	0.555	0.691	0.693	0.593	0.579	0.617	0.579	0.579	0.53	0.291	0.524	0.517	0.162	0.35	0.239							
r6960633		0.03	0.396	0.447	0.467	0.861	0.907		0.763	1.0	1.0	1.0	0.65	0.587	0.656	0.587	0.571	0.652	0.557	0.54	1.0	0.677	1.0	0.374	0.696	1.0	0.569	0.683	0.544	0.683	0.686	0.582	0.567	0.616	0.567	0.567	0.555	0.331	0.562	0.555	0.183	0.187	0.586	0.279							
r6731530		0.027	0.037	0.054	0.073	0.206	0.203	0.189		0.892	0.88	0.884	0.574	0.555	0.639	0.555	0.489	0.58	0.454	0.543	0.902	0.596	0.788	0.788	0.604	0.56	0.812	0.69	0.62	0.691	0.62	0.622	0.688	0.707	0.542	0.707	0.707	0.587	0.161	0.457	0.438	0.38	1.0	0.621	0.181						
r6402941		0.042	0.011	0.011	0.012	0.07	0.07	0.067	0.159		0.89	0.843	0.574	0.586	0.509	1.0	0.509	0.461	0.479	0.607	1.0	0.78	0.266	0.655	0.655	1.0	0.392	0.64	1.0	0.719	1.0	0.719	0.732	0.848	1.0	0.754	1.0	1.0	1.0	1.0	0.1	0.885	0.894	0.894	1.0	0.113	0.711	0.869			
r6968225		0.046	0.0030	0.0030	0.0	0.065	0.065	0.062	0.149	0.751		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.943	0.616	0.822	0.822	1.0	0.524	0.817	1.0	0.902	1.0	0.902	0.907	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.882	1.0	0.058	0.788	0.847				
r6968508		0.042	0.0070	0.0070	0.0010	0.07	0.07	0.067	0.156	0.71	0.949		1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.892	0.604	0.773	0.773	1.0	0.47	0.824	1.0	0.907	1.0	0.907	0.911	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.039	0.808	0.862						
r6950728		0.0	0.056	0.072	0.086	0.13	0.13	0.093	0.22	0.013	0.056	0.059		1.0	1.0	1.0	0.973	1.0	1.0	1.0	1.0	0.975	1.0	1.0	0.805	0.942	1.0	0.873	0.91	0.857	0.91	0.384	0.885	0.82	0.439	0.82	0.82	0.714	0.05	0.304	0.339	0.356	1.0	0.357	0.088						
r1992901		0.0	0.056	0.068	0.08	0.13	0.13	0.097	0.213	0.016	0.056	0.061	1.0		1.0	1.0	0.974	1.0	1.0	1.0	1.0	0.975	1.0	1.0	0.812	0.942	0.605	0.877	0.412	0.862	0.412	0.405	0.889	0.826	0.461	0.826	0.826	0.685	0.0020	0.261	0.298	0.322	1.0	0.338	0.032						
r6728141		0.0060	0.046	0.064	0.068	0.231	0.231	0.208	0.276	0.137	0.128	0.137	0.456	0.456		1.0	0.953	1.0	1.0	0.69	1.0	0.635	0.635	0.878	1.0	0.654	0.863	0.828	0.903	0.828	0.831	0.805	0.854	0.753	0.854	0.854	0.768	0.368	0.804	0.808	0.561	1.0	0.729	0.326							
r13245639		0.0	0.056	0.068	0.08	0.13	0.13	0.079	0.213	0.016	0.056	0.061	1.0	1.0	0.456		0.974	1.0	1.0	1.0	1.0	0.975	1.0	1.0	0.812	0.942	0.605	0.877	0.412	0.862	0.412	0.405	0.889	0.826	0.461	0.826	0.826	0.685	0.0020	0.261	0.298	0.322	1.0	0.338	0.032						
r6729302		0.0	0.06	0.072	0.086	0.125	0.125	0.074	0.169	0.012	0.052	0.057	0.922	0.924	0.404	0.924		0.947	0.972	0.946	1.0	0.947	1.0	1.0	0.743	0.881	0.544	0.821	0.436	0.845	0.436	0.423	0.876	0.76	0.76	0.614	0.031	0.266	0.285	0.25	0.804	0.319	0.056								
r6729068		0.0	0.056	0.075	0.089	0.13	0.13	0.095	0.226	0.014	0.056	0.059	1.0	0.976	0.456	0.976	0.898		1.0	0.973	1.0	0.975	1.0	1.0	0.806	0.943	1.0	0.873	0.934	0.857	0.394	0.387	0.855	0.821	0.443	0.821	0.821	0.718	0.035	0.286	0.322	0.358	1.0	0.351	0.07						
r12706860		0.0	0.039	0.05	0.075	0.127	0.127	0.078	0.156	0.025	0.062	0.067	0.907	0.909	0.503	0.909	0.832	0.887		1.0	0.974	1.0	1.0	0.829	0.938	0.689	0.731	0.354	0.875	0.354	0.339	0.787	0.677	0.677	0.558	0.093	0.205	0.22	0.374	1.0	0.297	0.072									
r7808659		0.0040	0.038	0.05	0.063	0.117	0.117	0.065	0.191	0.056	0.052	0.056	0.925	0.927	0.422	0.927	0.898	0.901	0.84		1.0	0.973	1.0	1.0	0.795	0.913	0.584	0.863	0.351	0.854	0.351	0.344	0.881	0.859	0.425	0.859	0.859	0.705	0.142	0.219	0.242	0.262	1.0	0.264	0.06						
r6402940		0.036	0.0010	0.0	0.0010	0.079	0.079	0.075	0.188	0.526	0.73	0.687	0.068	0.067	0.075	0.07	0.066	0.068	0.077	0.066		0.668	0.949	0.949	1.0	0.553	0.901	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.908	1.0	0.061	0.919	0.873					
r754284		0.0010	0.048	0.058	0.07	0.133	0.133	0.105	0.245	0.000	0.021	0.022	0.95	0.952	0.456	0.952	0.873	0.927	0.862	0.878		0.031	0.675	0.675	0.861	0.942	0.368	0.919	0.417	0.909	0.417	0.411	0.927	0.872	0.468	0.872	0.872	0.729	0.035	0.274	0.308	0.406	1.0	0.348	0.067						
r754281		0.05	0.0020	0.0030	0.0090	0.017	0.017	0.069	0.135	0.037	0.576	0.539	0.066	0.067	0.058	0.067	0.061	0.066	0.074	0.063	0.859	0.031	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.896	1.0	0.298	0.098	0.855					
r11768806		0.05	0.0020	0.0030	0.0090	0.017	0.017	0.069	0.135	0.037	0.576	0.539	0.066	0.067	0.058	0.067	0.061	0.066	0.074	0.063	0.859	0.031	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.896	1.0	0.298	0.098	0.855					
r6728142		0.015	0.0020	0.0060	0.0090	0.087	0.087	0.078	0.215	0.112	0.08	0.112	0.272	0.281	0.692	0.281	0.223	0.27	0.323	0.252	0.129	0.316	0.117	0.117	1.0	0.852	0.951	0.909	0.952	0.909	0.91	0.914	0.948	0.775	0.948	0.948	0.889	0.498	0.873	0.876	0.679	0.463	0.829	0.47							
r7801838		0.0020	0.037	0.053	0.068	0.131	0.133	0.103	0.199	0.0070	0.013	0.011	0.818	0.798	0.42	0.798	0.734	0.82	0.715	0.77	0.018	0.798	0.057	0.057	0.372	1.0	1.0	0.436	1.0	0.436	0.428	1.0	1.0	0.485	1.0	1.0	0.868	0.076	0.346	0.364	0.509	1.0	0.359	0.139							
r61874330		0.034	0.0	0.0	0.0010	0.031	0.031	0.079	0.165	0.321	0.517	0.532	0.072	0.028	0.067	0.028	0.021	0.075	0.04	0.025	0.739	0.01	0.868	0.868	0.099	0.066	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.889	0.912	0.824	1.0	0.635	0.845	0.752				
r6778754		0.0030	0.042	0.057	0.051	0.163	0.163	0.154	0.329	0.138	0.134	0.138	0.362	0.37	0.714	0.37	0.319	0.358	0.283	0.333	0.158	0.407	0.144	0.144	0.778	0.434	0.166	1.0	1.0	1.0	1.0	1.0	0.975	1.0	0.828	1.0	1.0	0.928	0.504	0.962	0.964	0.636	1.0	0.95	0.466						
r3757388		0.0020	0.066	0.074	0.043	0.1																																													

**Tabla suplementaria 1: Valores de D' y r<sup>2</sup> de los 44 polimorfismos analizados dos a dos.** En el triángulo inferior se muestran los valores de r<sup>2</sup>, en el superior los de D'.

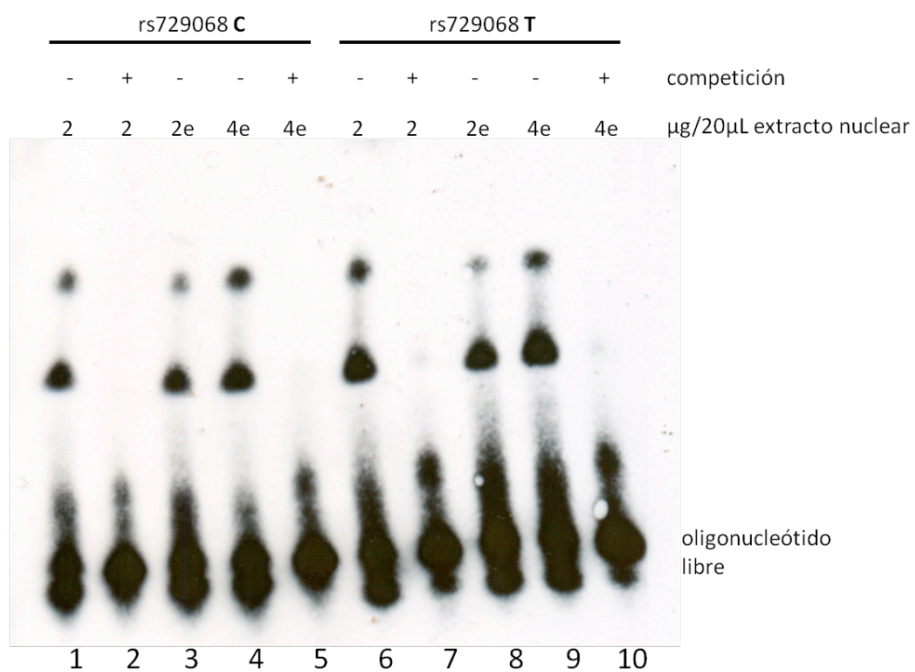




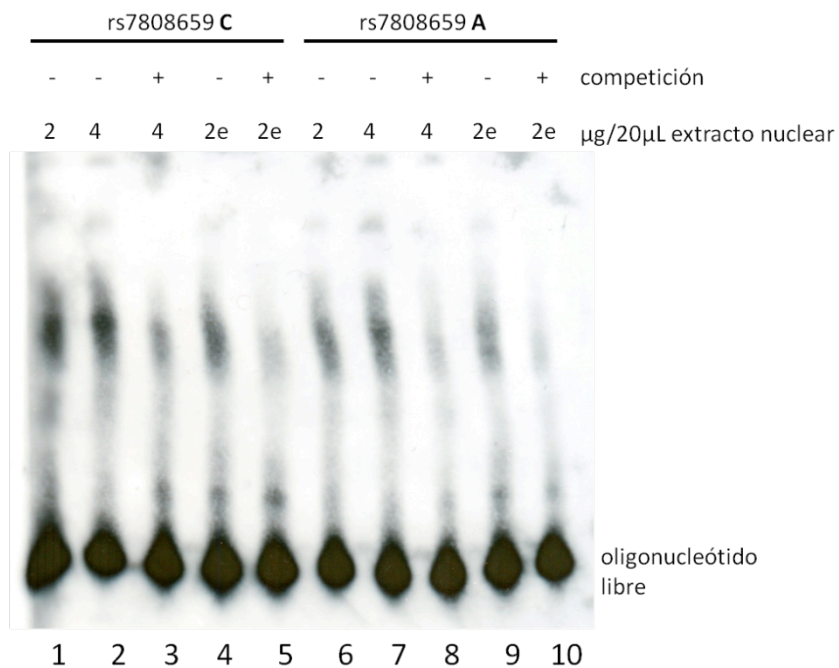
**Figura suplementaria 1: Ensayo EMSA de los dos alelos de rs6950728.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



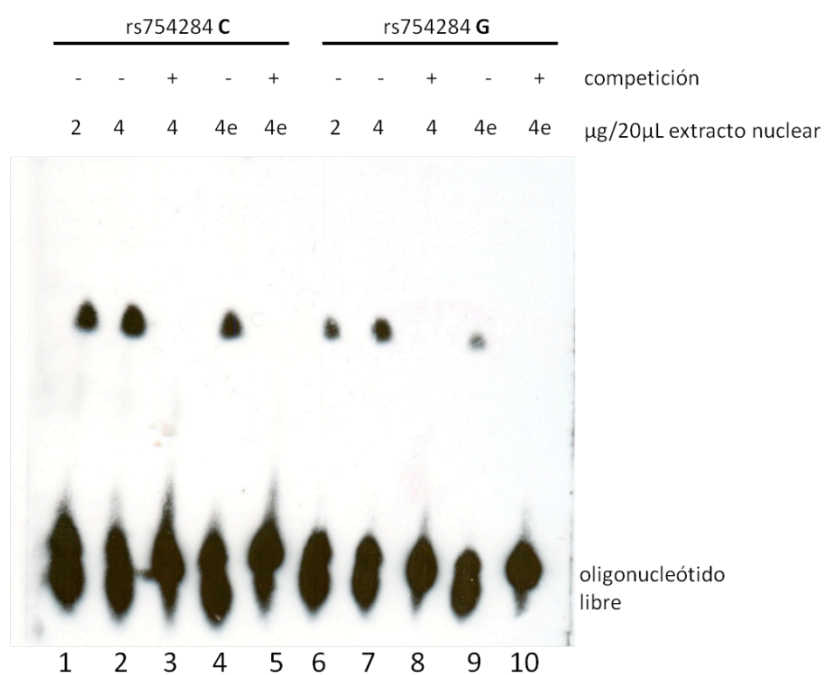
**Figura suplementaria 2: Ensayo EMSA de los dos alelos de rs11982901.** Ensayo EMSA 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



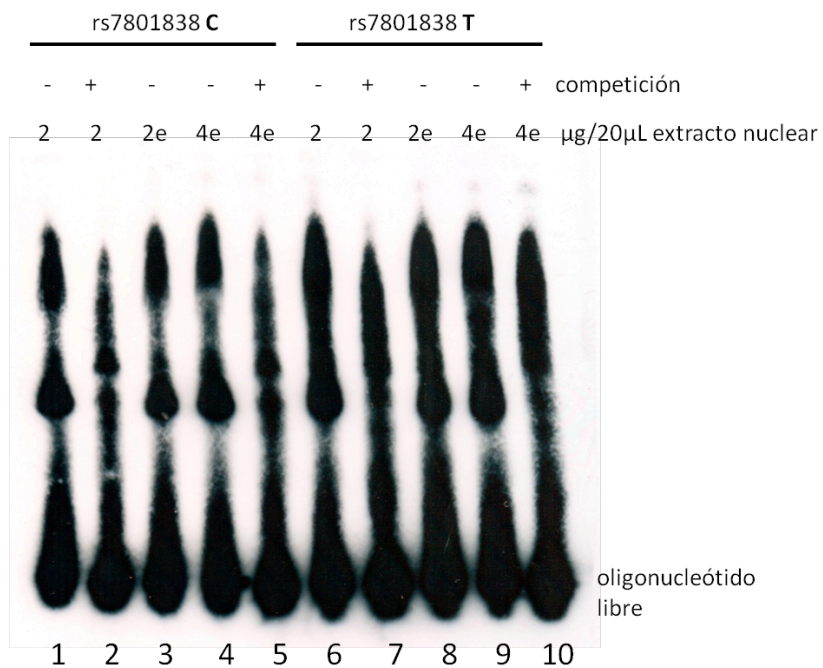
**Figura suplementaria 3: Ensayo EMSA de los dos alelos de rs729068.** Ensayo EMSA 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



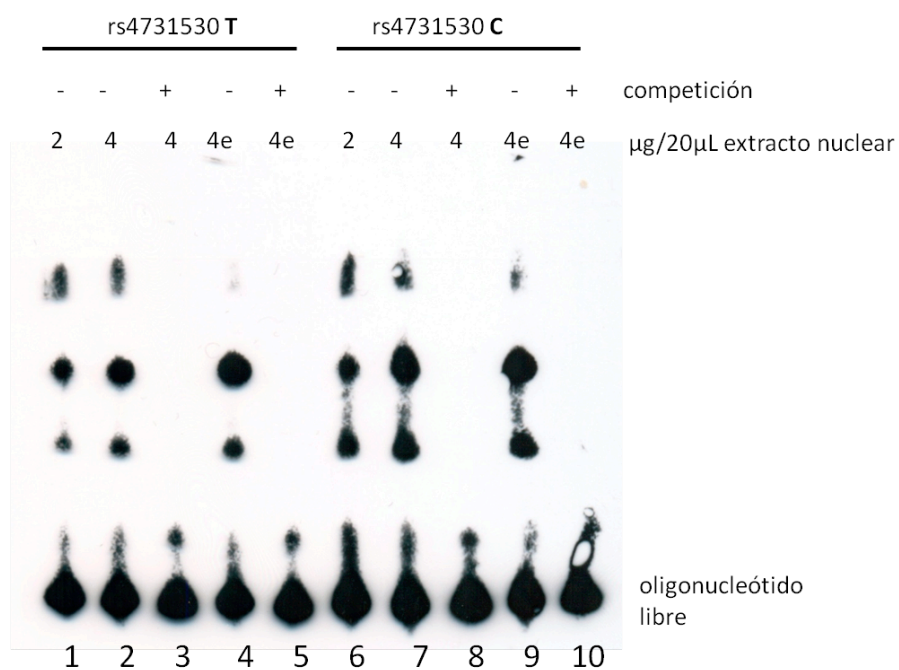
**Figura suplementaria 4: Ensayo EMSA de los dos alelos de rs7808659.** Ensayo EMSA con  $2\mu\text{g}$  y  $4\mu\text{g}$  de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



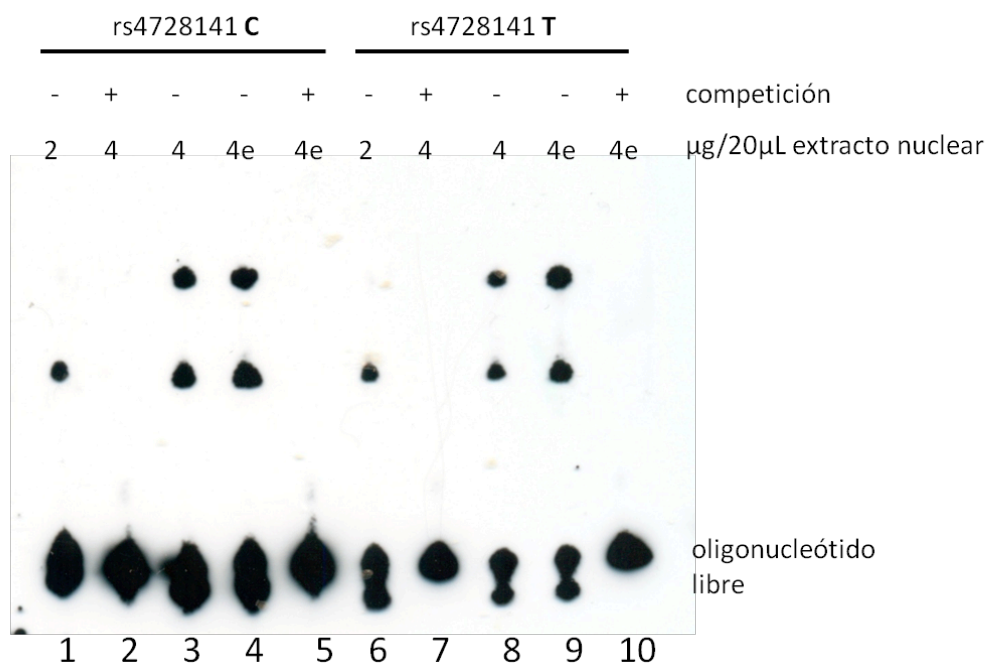
**Figura suplementaria 5: Ensayo EMSA de los dos alelos de rs754284.** Ensayo EMSA con 2µg y 4µg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



**Figura suplementaria 6: Ensayo EMSA de los dos alelos de rs701838.** Ensayo EMSA con 2 $\mu\text{g}$  y 4 $\mu\text{g}$  de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

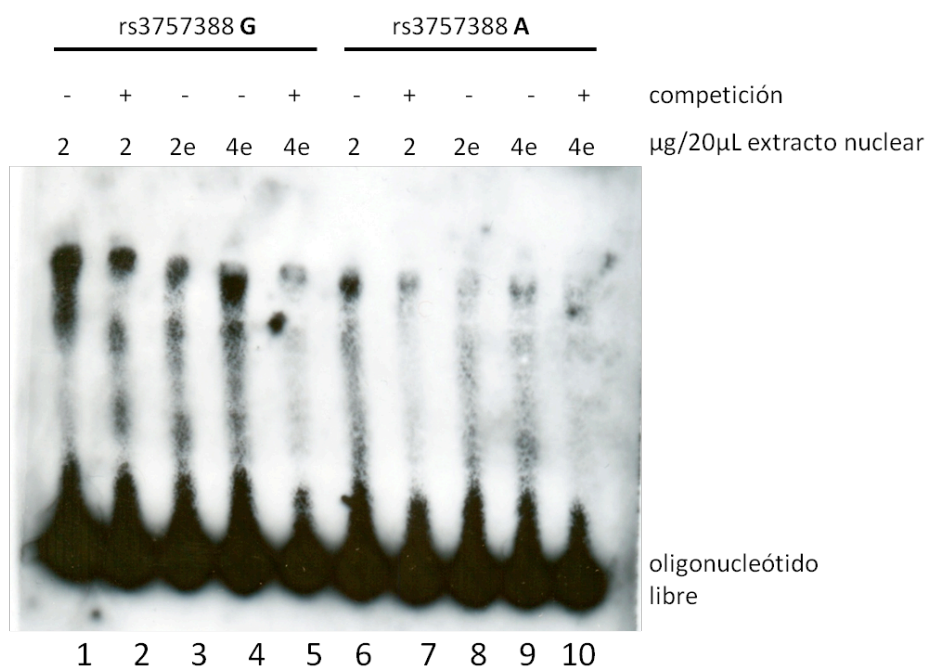


**Figura suplementaria 7: Ensayo EMSA de los dos alelos de rs4731530.** Ensayo EMSA con 2 $\mu\text{g}$  y 4 $\mu\text{g}$  de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

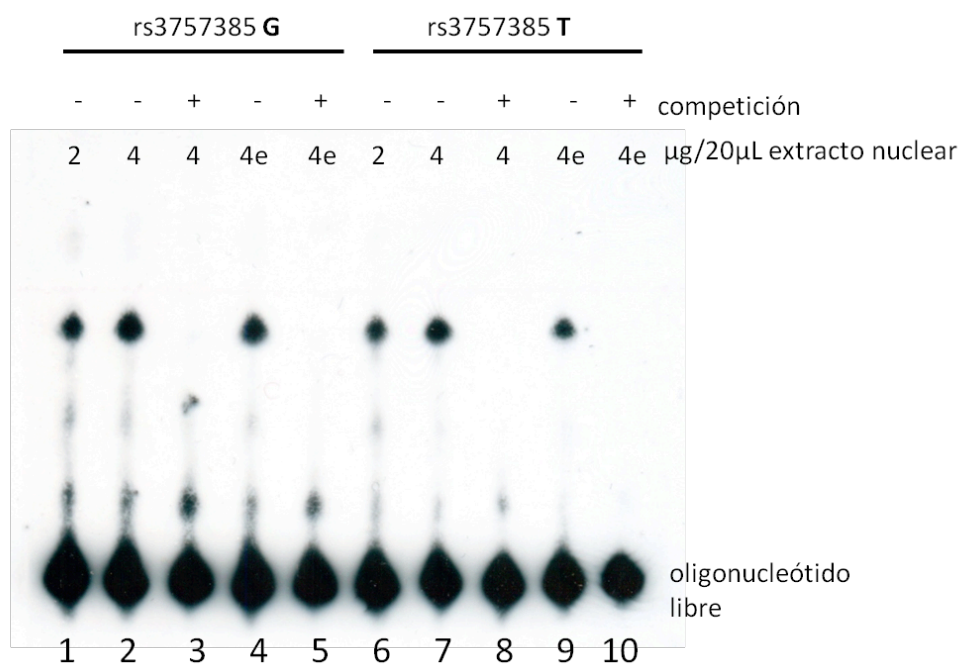


**Figura suplementaria 8: Ensayo EMSA de los dos alelos de rs4728141.** Ensayo EMSA con 2 $\mu\text{g}$  y 4 $\mu\text{g}$  de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

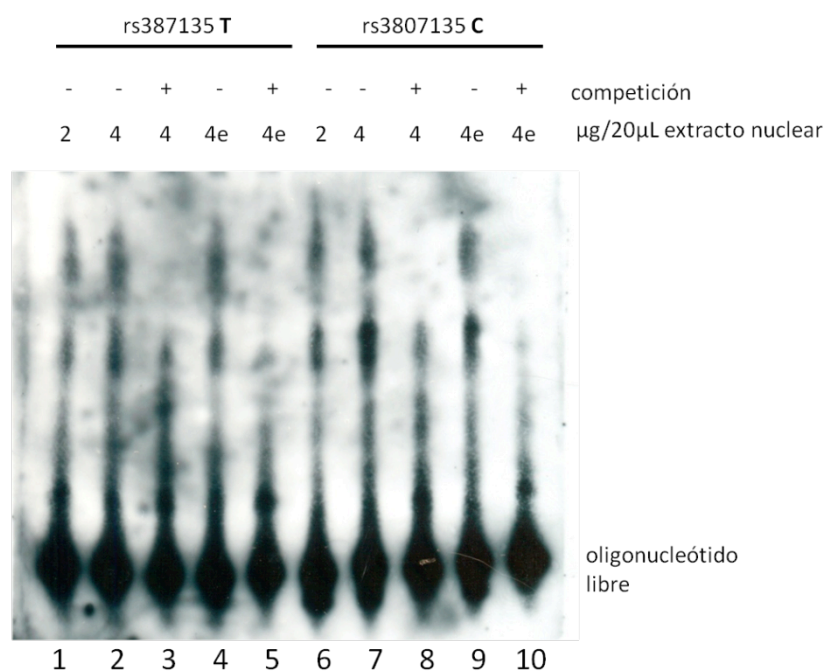




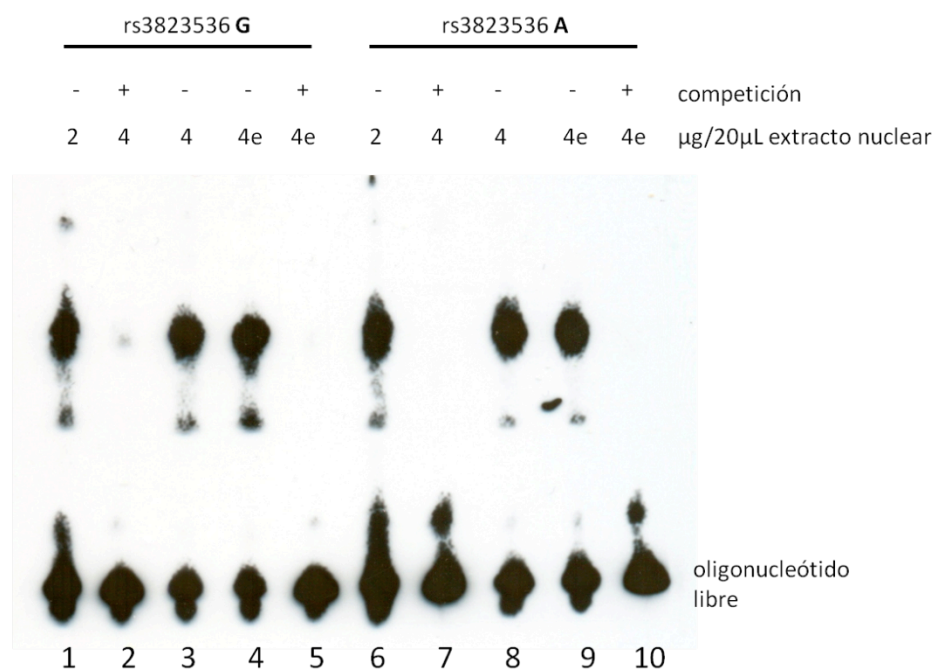
**Figura suplementaria 9: Ensayo EMSA de los dos alelos de rs3757388.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



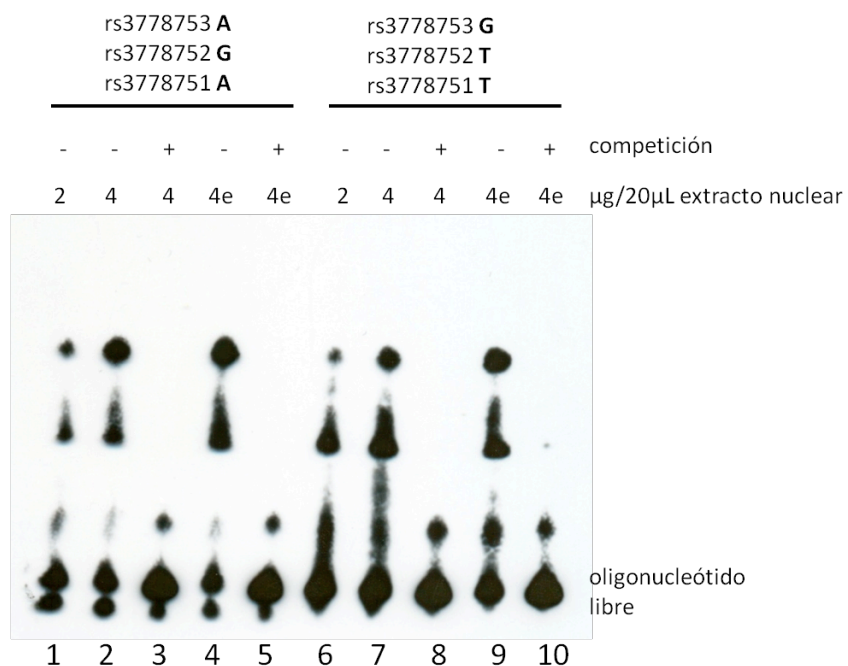
**Figura suplementaria 10: Ensayo EMSA de los dos alelos de rs3757385.** Ensayo EMSA con 2 $\mu\text{g}$  y 4 $\mu\text{g}$  de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



**Figura suplementaria 11: Ensayo EMSA de los dos alelos de rs387135.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



**Figura suplementaria 12: Ensayo EMSA de los dos alelos de rs3823536.** Ensayo EMSA con 2μg y 4μg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 2, 5, 7 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.



**Figura suplementaria 13: Ensayo EMSA de los dos alelos de los SNPs rs3778753, rs3778752 y rs3778751.** Ensayo EMSA con 2µg y 4µg de extracto nuclear de la línea celular WIL 2 NS en condiciones basales y tras estimulación (e). En los ensayos de competición (carriles 3, 5, 8 y 10) el oligonucleótido sin marcar estaba 200 veces más concentrado que el marcado.

## **DISCUSIÓN.**

---

## 1. Discusión.

El LES es una enfermedad autoinmune caracterizada por ser clínicamente muy heterogénea pudiendo afectar a múltiples sistemas orgánicos. Se han descrito 11 criterios clínicos de clasificación de la ACR, de los cuales un paciente con LES debe presentar al menos 4<sup>1-3</sup>. Además, el curso del LES se caracteriza por exacerbaciones y remisiones. Así, durante las fases agudas de la enfermedad, el paciente puede presentar diferentes manifestaciones clínicas. Esta clasificación y naturaleza del LES permite una gran heterogeneidad de fenotipos entre los pacientes dificultando así su diagnóstico, tratamiento y estudio.

Por otro lado, el LES también presenta una alta complejidad en su etiología en la que el componente genético juega un papel importante<sup>4-10</sup>. Varios estudios epidemiológicos, han observado que existen diferencias en las prevalencias de las manifestaciones clínicas del LES entre varios grupos étnicos y por lo tanto, con un *background* genético diferente<sup>11, 12</sup>. Sin embargo, más recientemente se ha demostrado que las diferencias genéticas presentes en una misma etnia también influyen el fenotipo clínico del LES<sup>13, 14</sup>. Este es el caso de Europa, en la cual se ha definido un fenómeno de subestructura mediante la utilización de marcadores informativos de ancestralidad (AIMs). Estos AIMs consisten en SNPs que presentan diferencias importantes en sus frecuencias alélicas entre poblaciones europeas de diferentes países<sup>15-18</sup>. Debido a estas evidencias previas, nos planteamos analizar el efecto que ejerce la subestructura europea sobre la prevalencia de las manifestaciones clínicas del LES en una amplia colección de 1413 pacientes europeos con LES procedentes de 9 países (**Publicación 1**). Para llevar a cabo este trabajo se genotiparon 6 SNPs descritos previamente como AIMs en relación a 11 criterios clínicos de la ACR y la fecha de inicio de la enfermedad. Como resultado, de las 12 variables clínicas analizadas, 3 de ellas (desorden inmunológico, úlceras orales y fotosensibilidad) mostraron resultados de asociación con alguno de los AIMs analizados. Por lo tanto, nuestro estudio confirma el efecto significativo que ejerce el fenómeno de subestructura presente en Europa en relación al fenotipo clínico del LES.

Una explicación de la complejidad de la etiología del LES viene dada por la alta heterogeneidad en su componente genético. A partir del año 2008 se han publicado 6 GWAS en LES que han dado a conocer más 30 variantes de susceptibilidad a la enfermedad<sup>15, 19-23</sup>. Sin embargo, todas estas señales no están asociadas en todas las etnias. Como sucede por ejemplo con los *loci* *PTPN22*, *PXK*, *FCGR2A* y los genes de la región HLA, que presentan resultados

discordantes dependiendo la etnia estudiada<sup>19-24</sup>. Estos resultados junto con la variabilidad genética existente dentro de Europa debido al fenómeno de subestructura, nos llevó a plantear la hipótesis de que estas diferencias en el tamaño del efecto de las señales de asociación al LES no sólo se vean entre etnias diferentes sino que también dentro de la misma etnia europea. De esta manera, nuestro siguiente trabajo consistió en el análisis de 25 *loci* de asociación con LES descritos previamente dentro de un estudio de asociación caso-control con 1742 pacientes con LES y 2101 controles de 17 colecciones procedentes de 10 países europeos (**Publicación 2**). Estas colecciones se dividieron en dos subgrupos, centro-europeos y sur-europeos, en base a los datos de los 6 AIMs analizados previamente (**Publicación 1**). En general, encontramos un sesgo hacia un mayor efecto del alelo de riesgo en las poblaciones del sur-europeas en relación a las centro-europeas confirmando así la heterogeneidad en el componente genético del LES a lo largo de Europa.

Como ya se ha comentado anteriormente, el LES se considerada una enfermedad genéticamente compleja ya que se han identificado alrededor de 30 variantes genéticas de susceptibilidad que confieren incrementos relativamente pequeños en el riesgo a padecerla, explicando así una modesta proporción de su heredabilidad. No todas estas variantes han sido confirmadas mediante estudios de replicación. Sin embargo, *IRF5* es uno de los genes asociados al LES con resultados ampliamente consolidados no sólo en LES sino que también en otras enfermedades autoinmunes como artritis reumatoide<sup>25</sup>, síndrome de Sjögren<sup>26, 27</sup>, esclerosis múltiple<sup>28, 29</sup>, enfermedad inflamatoria intestinal<sup>30</sup>, y cirrosis biliar primaria<sup>31, 32</sup>. Aún así la complejidad de los efectos descubiertos dentro de este gen no ha permitido todavía esclarecer su papel en la enfermedad. Por lo tanto, uno de nuestros objetivos fue analizar la variabilidad genética existente dentro del gen *IRF5* en relación a la susceptibilidad al LES. Para ello llevamos a cabo un estudio de asociación caso-control además de un análisis de su distribución haplotípica con 9 *tag* SNPs que cubrían la variabilidad genética de *IRF5* (**Publicación 3**). En este mismo estudio también incluimos todos los polimorfismos de *IRF5* con una repercusión funcional conocidos hasta la fecha. Estos polimorfismos consistieron en el SNP rs2004640 cuyo alelo T crea un *donor splice site* que promoverá la expresión de la variante exónica 1B<sup>33</sup>; el indel del exón 6 que consiste en una inserción/delección de 30 pb presente dentro del exón 6<sup>34-36</sup>; y por último, el SNP rs10954213 localizado en una señal de poly A dentro de la zona 3'-UTR del gen cuyo alelo A da lugar a un mRNA de menor tamaño y mayor expresión<sup>34, 35, 37, 38</sup>. Como conclusión general, nuestros resultados sugieren la existencia de dos efectos independientes y opuestos dentro del gen *IRF5*. Un efecto protector localizado en la región promotora del gen junto con un efecto de susceptibilidad marcado por una epístasis entre los tres polimorfismos funcionales anteriormente descritos.



Como se ha referido anteriormente, *IRF5* presenta varios polimorfismos con diferentes repercusiones funcionales. Uno de ellos consiste en un indel localizado en el promotor proximal del gen cuya unidad de repetición consiste en una secuencia de 5 pb (CGGGG) que se repite 4 veces (inserción) o 3 veces (delección)<sup>34</sup>. Este polimorfismo está asociado al LES y a enfermedad inflamatoria intestinal<sup>30, 39</sup>. Es importante destacar que dentro de todos los polimorfismos funcionales de *IRF5*, el SNP rs10954213 (poly A), el SNP rs2004640 (*splice donor site*) y el nuevo CGGGG indel se han caracterizado como reguladores de su expresión génica. Sin embargo, sus mecanismos individuales permanecen aún por detallar.

Además, cabe la posibilidad de la existencia de una nueva repercusión funcional en relación a la regulación de su expresión génica ya que el efecto protector encontrado en nuestro trabajo previo está localizado en la región promotora del gen. Debido a estas evidencias previas, analizamos los datos de expresión de *IRF5* obtenidos a partir de 4 estudios realizados con *microarrays* de expresión<sup>40-43</sup> en relación a los polimorfismos funcionales y a un amplio número de polimorfismos presentes en el gen *IRF5* y su región promotora sin repercusión funcional conocida (**Publicación 4**). De todos los polimorfismos funcionales analizados, el SNP rs10954213 (poly A) mostró la mayor contribución a la regulación de la expresión génica. Además, también confirmamos que el haplotipo de riesgo #6 descrito en nuestro anterior estudio (**Publicación 3**) está asociado con una mayor expresión de *IRF5*. Sin embargo, el efecto protector no se asoció completamente con una disminución de la expresión del gen, lo que nos indica que la explicación de la asociación de *IRF5* con el LES incluye la identificación de más polimorfismos funcionales además de otro tipo de modificaciones funcionales no relacionados con la regulación de la expresión.

Considerando como punto de partida la búsqueda de nuevos polimorfismos funcionales que pudieran explicar el efecto protector de *IRF5* en relación al LES, nos planteamos el estudio detallado de la región promotora del gen. El primer polimorfismo clave en la búsqueda fue el SNP rs729302 el cual define los haplotipos protectores #1 y #2 descritos en nuestro estudio previo (**Publicación 3**). Además, este SNP rs729302 está localizado en la región promotora del gen, por lo tanto, podría así influir en la unión de factores de transcripción y con ello en la regulación de la expresión génica. Sin embargo, este SNP rs729302 no presentó cambios en la unión de factores de transcripción mediante ensayos EMSA. Por lo tanto, nos centramos en la región promotora de *IRF5* ya que consideramos que el SNP protector rs729302 funcionaría como *tag* SNP de otro nuevo polimorfismo con una repercusión funcional que explicaría el efecto protector de *IRF5* en relación al LES localizado en esa zona. Para llevar a cabo este objetivo realizamos un mapa de LD con un total de 54 polimorfismos presentes en *IRF5* y su región promotora. A partir de estos resultados, seleccionamos varios polimorfismos en base a su información de  $r^2$  con el SNP protector rs729302 y al estudio de la distribución haplotípica. En

esta fase de búsqueda, solamente encontramos un polimorfismo con repercusión funcional validada mediante ensayos EMSA y gen *reporter* de luciferasa localizado en la región promotora de *IRF5*: SNP rs13245639.

Como parte de la búsqueda de polimorfismos funcionales dentro de la región promotora de *IRF5*, decidimos confirmar el papel funcional del CGGGG indel previamente descrito<sup>30, 39, 44</sup>. Estudio previos mediante ensayos EMSA mostraron que la inserción 4xCGGGG crea un lugar de unión adicional en la región promotora de *IRF5* para el factor de transcripción SP1<sup>30, 39, 44</sup>. Además, ensayos *reporter* con mini-genes y análisis *immunoblot* muestran que las células homocigotas para la inserción 4xCGGGG presentan un mayor nivel de expresión de *IRF5* en comparación a las células homocigotas para la delección 3xCGGGG<sup>39</sup>. Nuestros resultados confirman la presencia de unión alélica diferencial mediante ensayos EMSA, además de una expresión alélica diferencial del gen de la luciferasa mediante ensayos gen *reporter*.

Otra fuente de información utilizada para la búsqueda de polimorfismos funcionales consistió en el análisis de los datos de expresión de *IRF5* obtenidos a partir de la secuenciación de alto rendimiento (NGS, *next generation sequencing*) del transcriptoma mediante la tecnología RNA-seq<sup>45</sup>. Estos resultados nos proporcionan una mayor precisión en la evaluación de la expresión génica en comparación con los estudios basados en *microarrays*<sup>46</sup>. Otra implementación importante en este estudio es la utilización de la base de datos de genotipado *1000 Genomes*. Esta base de datos está basada en el genotipado mediante secuenciación del genoma completo con tecnología de secuenciación de alto rendimiento. Por lo tanto, el número de polimorfismos incluidos en esta base de datos es mucho mayor en comparación con el Proyecto Internacional HapMap proporcionándonos la posibilidad de descubrir nuevos polimorfismos con una repercusión funcional a nivel de la regulación de la expresión génica localizados en *IRF5* o su región promotora y que puedan así explicar el papel protector de este gen frente al LES. A partir de este análisis encontramos un nuevo polimorfismo funcional que consiste en un indel de 14 pb localizado en la región promotora de *IRF5*: rs11269962. Este indel rs11269962 forma parte del mejor modelo de regresión lineal para la explicación de los datos de expresión de *IRF5* procedentes de RNA-seq junto con el SNP rs11973360 presente en el gen de la *TNPO3*. Concretamente, su alelo delección promueve un aumento de la expresión de *IRF5* con una contribución significativa en el modelo. A partir de este resultado, nos planteamos testar funcionalmente el indel rs11269962 como buen candidato para la explicación de la protección del gen *IRF5* al LES. Este indel presentó unión alélica diferencial mediante ensayos EMSA además de una expresión alélica diferencial del gen de la luciferasa mediante ensayos gen *reporter*.

Finalmente, la búsqueda de polimorfismos funcionales mediante el estudio del mapa de LD de *IRF5*, resultó en la validación funcional de 3 polimorfismos: rs13245639, CGGGG indel y el indel rs11269962 (todos ellos localizados en la región promotora del gen). Estos polimorfismos podrían explicar funcionalmente el efecto protector de *IRF5* frente al LES. Por lo tanto, es necesario el estudio y análisis de su asociación y distribución haplotípica en relación con la enfermedad. Teniendo como base este objetivo, realizamos un estudio de asociación caso-control con un total de 2092 pacientes con LES y 1729 controles. En este estudio están incluidos los 3 polimorfismos funcionales previamente descritos (rs13245639, CGGGG indel y el indel rs11269962) junto con los 9 *tag* SNPs de *IRF5* analizados en nuestro trabajo previo (**Publicación 3**). Como resultados generales, todos los polimorfismos funcionales analizados están asociados de manera significativa con la susceptibilidad al LES en nuestra colección de muestras. Seguidamente, gracias al estudio de la nueva distribución haplotípica, podemos decir que se mantienen los dos efectos independientes y opuestos de susceptibilidad y protección previamente descritos (**Publicación 3**). En concreto, el efecto protector agrupa en sus haplotipos de manera específica, los alelos de las variables funcionales que disminuyen la expresión del gen (alelo T del SNP rs13245639, alelo delección CGGGG indel y alelo inserción del indel rs11269962). Por el contrario, el haplotipo de susceptibilidad agrupa la combinación alélica opuesta que promueve el aumento de la expresión de *IRF5* pero en este caso de manera no específica (alelo C del SNP rs13245639, alelo inserción CGGGG indel y alelo delección del indel rs11269962). Estos resultados apoyarían los resultados previos encontrados en el estudio de la expresión de *IRF5* con datos de *microarrays* (**Publicación 4**).

### **1.1. Asociación de las manifestaciones clínicas del LES con la subestructura genética presente en la población europea: Publicación 1.**

El fenómeno de subestructura se puede definir como la estratificación o subdivisión de una etnia en base a sus diferencias en el *background* genético. Varios estudios han demostrado mediante el genotipado de múltiples polimorfismos en colecciones de etnia europea la existencia de una subestructura genética clara<sup>16-18, 47</sup>. Mediante el análisis de componentes principales de estos genotipos, se observó que existían dos componentes principales que definen los ejes geográficos norte-sur y este-oeste de Europa. En todos ellos, la primera componente principal consiste en el eje norte-sur que explicaría la mayor fuente de variabilidad genética dentro de la etnia europea. La segunda componente principal, independiente y acumulativa de forma aditiva a la primera, consiste en el eje este-oeste dentro de la etnia europea<sup>16-18, 47</sup>. Los SNPs que mejor definen las dos componentes principales se consideran marcadores informativos de ancestralidad (AIMs). Estos AIMs se caracterizan por mostrar grandes diferencias en sus frecuencias alélicas entre los subgrupos europeos<sup>15-18</sup>.

Concretamente, la selección de nuestros AIMs se basó en uno de estos estudios en el que se genotiparon más de 300000 SNPs en una colección de más de 10000 individuos de etnia europea<sup>17</sup>. Concretamente seleccionamos los SNPs rs6730157 y el rs382259 ya son aquellos que presentan una contribución mayor en la definición de la primera componente principal (eje norte-sur). Por otro lado, también seleccionamos los SNPs rs12203592 y el rs354690 ya que son aquellos que presentan una contribución mayor en la definición de la segunda componente principal (eje este-oeste). Además, el SNP rs12203592 ha mostrado diferencias en sus frecuencias alélicas correlacionadas significativamente con la determinación del color de ojos, pelo, piel y la capacidad de la piel para asimilar la radiación solar en población europea<sup>15</sup>. Todos estos fenotipos tienen una distribución geográfica conocida a nivel europeo aportándole la capacidad de AIM a este SNP rs12203592. Del mismo modo, también genotipamos como posible AIM el SNP rs12913832, analizado en el mismo estudio previo, el cual está significativamente correlacionado con la determinación del color pelo y la capacidad de la piel para asimilar la radiación solar en población europea<sup>15</sup>. Por último, el SNP rs4988235 se encontró asociado con la persistencia de la actividad de la enzima lactasa en la edad adulta<sup>48</sup>. Este SNP rs4988235 presenta diferencias en sus frecuencias alélicas a lo largo de Europa, pudiendo así funcionar como un AIM.

Estos AIMs utilizados en nuestro estudio fueron suficientes para mostrar el fenómeno de subestructura europea en nuestras colecciones. La variación más importante consistió en el eje norte-sur al igual que los resultados observados en estudios previos<sup>16-18, 47</sup>. De este modo, la correlación más importante fue la existente entre la primera componente principal y la latitud ( $r = 0,47$ ). Además, los AIMs que mostraron mayores diferencias de sus frecuencias alélicas a lo largo de nuestras colecciones europeas (rs6730157, rs382259 y el rs12913832), fueron también los que presentaron una mayor correlación con la latitud ( $r > 0,6$  en todos ellos). Este resultado es importante teniendo en cuenta que en nuestro estudio no están incluidas colecciones de muestras de países europeos con localizaciones geográficas más extremas como podrían ser Escandinavia, Inglaterra o Rusia que aumentarían el rango de frecuencias de los AIMs mejorando sus correlaciones con los ejes geográficos<sup>16-18</sup>.

Por otro lado, los AIMs utilizados en nuestro estudio presentan una limitación importante ya que no son suficientes para clasificar a los sujetos de manera individual. Sin embargo, consideramos que son suficientes para detectar una importante fracción de la heterogeneidad clínica del LES asociada al fenómeno de subestructura europea. De hecho, dos de nuestros resultados se encontraron asociados con más de un AIM indicando un cierto nivel de redundancia a pesar de la inexistente correlación, a nivel de LD, que existe entre sus genotipos.

De este modo, estos AIMs serían no sólo suficientes sino que también efectivos para el objetivo de este trabajo consistente en estudiar la variabilidad del fenotipo clínico del LES al largo de la subestructura europea.

El primer resultado de este trabajo consistió en la asociación significativa del AIM SNP rs382259 y la primera componente principal con el desorden inmunológico. Este desorden o alteración inmunológica se basa en la producción de autoanticuerpos anti-dsDNA, anti-Smith, y anti-fosfolípido (concretamente anti-cardiolipina y/o anticoagulante lúpico). Interpretamos esta asociación como un aumento en la prevalencia del desorden inmunológico en las poblaciones del sur de Europa. Es importante destacar, que nuestros resultados confirman la misma asociación encontrada en estudios previos realizados con colecciones de muestras con ascendencia euro-americana<sup>13, 14</sup>. En ambos estudios observaron un aumento significativo de la prevalencia del desorden inmunológico en las poblaciones del sur de Europa. Además, en uno de ellos también se observó un aumento significativo de la prevalencia del desorden inmunológico en las poblaciones del oeste de Europa<sup>14</sup>. Interesantemente, en ambos estudios se encontró un aumento significativo de la producción de autoanticuerpos anti-dsDNA y anti-cardiolipina en las poblaciones del sur de Europa<sup>13, 14</sup>.

La explicación de estas diferencias en las prevalencias del desorden inmunológico en general aún está por detallar. En primer lugar, se conocen diferencias en la producción de autoanticuerpos individuales entre distintas etnias<sup>49-55</sup>. Por ejemplo, la etnia africana presenta un aumento significativo en la producción de autoanticuerpos anti-Smith y anti-RNP comparado con la etnia europea<sup>50-52, 55</sup>. Sin embargo, no existe ninguna evidencia previa de un aumento generalizado en la producción de autoanticuerpos característicos del LES. Incluso en algunos estudios se ha visto que una misma etnia presenta aumento en la producción de unos autoanticuerpos concretos y disminución de otros<sup>53-55</sup>. Por lo tanto, parece que el *background* genético está influenciando respuestas inmunológicas específicas, en lugar de defectos generalizados que darían lugar a una autoinmunidad mediada por autoanticuerpos. Un posible ejemplo de esta hipótesis vendría dado por los genes localizados en la región HLA los cuales también sufren diferencias en sus frecuencias alélicas a lo largo de Europa. Por ejemplo, las frecuencias alélicas de los alelos de riesgo al LES (HLA-DRB1\*0301 y HLA-DRB1\*1501) presenta una clina geográfica noroeste-sureste en Inglaterra<sup>56, 57</sup>. Al mismo tiempo, los alelos de HLA de clase II también se asociaron con la prevalencia de autoanticuerpos anti-dsDNA<sup>58</sup> y anti-cardiolipina<sup>59, 60</sup>. Por lo tanto, las variaciones en las frecuencias alélicas de los genes de la región HLA a lo largo de la subestructura europea, podrían estar influenciado las diferencias en la producción de autoanticuerpos y prevalencia del desorden inmunológico.

En relación a nuestro segundo resultado, su interpretación es más confusa. Encontramos una asociación con la prevalencia a padecer úlceras orales con dos de los AIMs incluidos en nuestro estudio: SNP rs12913832 y SNP rs382259. Sin embargo, ambas asociaciones se pierden cuando tenemos en cuenta la localización del centro de recogida como covariable en el análisis. Especialmente debido a que las colecciones procedentes de España y Portugal de nuestra colección presentan una alta prevalencia a úlceras orales comparado con las demás poblaciones del estudio. De todos modos, este análisis no invalidaría nuestros resultados ya que teniendo en cuenta los genotipos de los AIMs asociados, podemos decir que el aumento de prevalencia de úlceras orales se encontraría localizado en las poblaciones de sur-oeste europeo al igual que España y Portugal. En referencia a los dos estudios previos con población de ascendencia euro-americana, ambos presentaron una tendencia hacia al aumento de la prevalencia a padecer úlceras orales en el sur de Europa<sup>13, 14</sup>. Sin embargo, ambos resultados presentaron una asociación muy débil que se perdía cuando ajustan el análisis con covariables como sexo, tiempo de seguimiento y, como en nuestro caso, colección. Un punto importante de nuestros resultados consiste en la validez de las claras diferencias de nuestros datos de prevalencia de úlceras orales en comparación con otros estudios. Concretamente, se ha descrito que la prevalencia de úlceras orales, en dos colecciones independientes de unos 500 pacientes con LES españoles, ronda los valores 46,4%-54,3% muy comparables con nuestros datos de colecciones españolas con un 46,7 % de prevalencia<sup>61</sup>. Además, nuestro dato 49,1% de prevalencia de úlceras orales de Portugal, también sería comparable con el 45% descrito en un estudio previo con 544 pacientes portugueses con LES<sup>62</sup>. Por último, y en el otro extremo que validaría nuestras diferencias, un estudio previo muestra un dato de prevalencia de úlceras orales de 12,5 % también comparable con nuestro dato del resto de colecciones europeas de 18,5 %<sup>63</sup>.

Por último, el tercer resultado consistió en la asociación de la prevalencia de padecer fotosensibilidad con dos de los AIMs incluidos en nuestro estudio: SNP rs12913832 y el SNP rs12203592. La asociación del AIM rs12913832 no se plasmó en el manuscrito de este trabajo debido a que su nivel de asociación no llegaba al umbral impuesto por la corrección de Bonferroni aplicada ( $p = 0,008$ ). De todos modos, la asociación estaría en el valor de significación límite para poder ser considerada. Concretamente, el alelo G del AIM rs12913832 está asociado con un aumento de la prevalencia de la fotosensibilidad ( $p = 0,009$ ). Mediante este resultado podemos decir que existe un aumento de prevalencia de la fotosensibilidad en el norte de Europa. Además, la asociación de este AIM rs12913832 con la prevalencia de la fotosensibilidad es independiente del posible efecto de cualquier otro AIM incluido en este estudio. Por otro lado, la asociación del aumento de la prevalencia de la fotosensibilidad con el alelo T del AIM rs12203592 tiene una interpretación geográfica más confusa debido a la alta

dispersión de sus frecuencias alélicas a lo largo de nuestras colecciones sin seguir ningún eje geográfico concreto.

Este mismo aumento de la prevalencia de la fotosensibilidad en el norte de Europa, ya se había descrito en ambos estudios previos con poblaciones de euro-americanas<sup>13, 14</sup>. Este efecto de la subestructura europea en la prevalencia de fotosensibilidad, tendría una posible hipótesis explicativa ya que los pacientes con LES del norte de Europa, por lo general, tienen la pigmentación de la piel más clara lo cual les aportará una mayor sensibilidad a al daño producido por la luz solar<sup>64, 65</sup>. Además, esta hipótesis tendría más peso ya que ambos AIMs asociados, rs12913832 y rs12203592, están previamente involucrados en la determinación del color de ojos, pelo, piel y capacidad de la piel para asimilar la radiación solar en población europea<sup>15</sup>. Por lo tanto, esta hipótesis necesita estudios posteriores que la corroboren.

Finalmente, nuestro estudio refuerza la evidencia de una relación entre la variabilidad del fenotipo clínico del LES con la subestructura europea. Estudios futuros que detallen este mecanismo podrían entender la heterogeneidad clínica presente en el LES. Además, estos hallazgos también serán importantes para definir las relaciones individuales entre *loci* de asociación al LES y determinadas manifestaciones clínicas.

## **1.2. Existencia de un sesgo en el tamaño de los efectos de múltiples *loci* de asociación al LES a lo largo de Europa: Publicación 2.**

Como ya se ha referido anteriormente, es bien conocida la existencia de diferencias de tamaños del efecto de múltiples *loci* de riesgo al LES dependientes del *background* genético. Una de las diferencias más consolidadas, consiste en el aumento del número de alelos de riesgo presente en individuos con ascendencia amerindia comparado con otras etnias<sup>66</sup>. Sin embargo, la hipótesis planteada en este estudio consistió en buscar estas mismas diferencias en el tamaño del efecto de los *loci* de riesgo al LES pero dentro de la misma etnia europea influenciada por el previamente descrito fenómeno de subestructura genética<sup>16-18, 47</sup>. Para el planteamiento de esta hipótesis, nos basamos en la existencia de clinas europeas de frecuencias alélicas de varios *loci* de riesgo al LES como son: *PDCDI*<sup>67</sup>, genes de la región HLA<sup>56, 57</sup> y *PTPN22*<sup>68</sup>. Por otro lado, también consideramos la reciente existencia de diferencias en la prevalencia de la manifestaciones clínicas del LES a lo largo de la subestructura europea<sup>13, 14</sup> (**Publicación 1**). Es importante destacar que todas estas clinas correlacionan con el eje norte-sur de Europa considerado el más explicativo dentro de la subestructura europea<sup>16-18</sup>.

En primer lugar, debemos destacar que la subdivisión geográfica realizada en este estudio es coherente con la distribución de frecuencias alélicas de 3 de los 6 AIMs incluidos en nuestro

estudio previo (**Publicación 1**). Para ello construimos un score norte-sur (score N/S) teniendo en cuenta la distribución de las frecuencias alélicas de los AIMS: rs6730157, rs382259 y el rs12913832. Todos ellos han mostrado evidencias claras de su comportamiento como AIMS en estudio previos contribuyendo principalmente a la distinción entre el eje norte-sur europeo<sup>15, 17</sup>. La selección de estos tres AIMS para la creación del score N/S viene promovida porque son aquellos que presentan un mayor rango de datos de frecuencias alélicas a lo largo de nuestras colecciones distribuyéndolas claramente en el eje norte-sur europeo, además de presentar todos ellos efectos independiente con un valor de  $r^2$  entre ellos  $< 0,03$ . Así en la **tabla 1 de la Publicación 2**, podemos observar que las colecciones consideradas como centro-europeas serán aquellas que presente un valor del score N/S  $> 54,9$  (Holanda, Bélgica, Alemania, República Checa, Eslovaquia y Hungría). Por último, las poblaciones del sur de Europa muestran un salto en el dato del score N/S, presentando todas ellas valores de score N/S  $< 47,6$ : Portugal, España, Italia y Grecia.

Como resultado general de nuestro estudio encontramos un sesgo significativo que describe mayores tamaños de efecto de múltiples *loci* de riesgo al LES en poblaciones del sur comparado con las poblaciones del centro de Europa. Además, también encontramos una menor frecuencia de los alelos de riesgo de estos *loci* tanto en pacientes con LES como en controles del sur comparado con el centro de Europa.

Varios análisis verifican y validan la existencia del sesgo hacia mayores tamaños de efecto en las poblaciones del sur de Europa. De manera descriptiva, del total de las 20 señales de asociación al LES analizadas en nuestro estudio, 15 de ellas mostraron un aumento de su tamaño de efecto en las poblaciones del sur comparado con las del centro de Europa. Esta distribución es significativamente diferente de la distribución binomial de las OR que resultaría de manera aleatoria ( $p = 0,019$ ). Este resultado indicaría un sesgo generalizado claro hacia mayores tamaños de efecto en las poblaciones del sur de Europa. Además, la magnitud de esta diferencia es significativa mostrando una mayor media geométrica de las OR del sur ( $G_{media} = 1,46 \pm 1,30$ ) comparada con la del centro de Europa ( $G_{media} = 1,34 \pm 1,17$ ;  $p = 0,02$ ). El análisis de estos datos se realizó teniendo en cuenta que las 20 señales de asociación incluidas son independientes, ya que parte del efecto atribuido a cada una de ellas puede ser debido a la dependencia con otras de las señales incluidas en el estudio. Para testar esta independencia, recopilamos los datos de OR para cada uno de las 20 señales de asociación obtenidas a partir de un análisis de regresión logística condicionada con todas las otras señales de asociación incluidas como covariables. Del mismo modo, obtenemos una diferencia significativa comparando la media geométrica de las OR condicionadas de las 20 señales de asociación entre el sur y el centro de Europa (sur:  $G_{media} = 1,43 \pm 1,24$  vs. norte:  $G_{media} = 1,28 \pm 1,18$ ;  $p =$



0,02). Por lo tanto, la concordancia de estos resultados nos aporta la validez necesaria para comparar cada uno de los efectos de manera independiente.

Otra de las ventajas de esta comparación es que en todo momento la obtención de los datos de OR se realizó con un meta-análisis mediante la aproximación Mantel-Haenszel la cual tiene en cuenta la posible heterogeneidad de efectos entre múltiples colecciones. Sin embargo, una limitación de estos primeros resultados consiste en que no nos informan de la naturaleza de estas diferencias del tamaño del efecto entre el centro y sur de Europa.

En un esfuerzo hacia la búsqueda de la naturaleza de la diferencia de tamaños de efectos entre el centro y sur de Europa, realizamos una serie de análisis adicionales. El primero de ellos fue el conteo del número de alelos de riesgo por muestra. Este análisis es más informativo, pero no nos aporta una información global por colección. Aún así, estratificando nuestras muestras en centro y sur de Europa, además de entre casos con LES y controles, obtuvimos que el número de alelos de riesgo aumenta en este orden: controles del sur de Europa, controles del centro de Europa, casos con LES del sur de Europa y casos con LES del norte de Europa. En base a estos resultados, aplicamos especial atención a la comparación entre casos y controles del centro y del sur de Europa. Concretamente, la diferencia más significativa se encontró entre los controles de ambas procedencias ( $p = 1,1 \times 10^{-8}$ ), siendo mucho menos significativa la diferencia entre los pacientes con LES ( $p = 0,016$ ). También realizamos un análisis dependiente del anterior en el que la suma de alelos de riesgo está ponderada por la OR de cada señal en cada una de los grupos del centro y sur de Europa obtenidas mediante el test Mantel-Haenszel (GRS, *genetic risk score*). Estos resultados refuerzan nuestras evidencias previas ya que observamos un aumento significativo del GRS en las poblaciones del sur en comparación con las poblaciones del centro de Europa ( $p = 1,8 \times 10^{-32}$ ).

Por lo tanto, parece que la naturaleza de esta desviación hacia un aumento del tamaño del efecto en poblaciones del sur de Europa es debido a una disminución generalizada en el número de alelos de riesgo en sus controles en relación a los controles del centro. Sin embargo, esta misma diferencia en ambos grupos de pacientes con LES es mucho menos importante. Consideramos que esto es debido a la naturaleza del LES como enfermedad genéticamente compleja con un modelo de herencia poligénica. Este modelo de herencia poligénica se produce en enfermedades, que como el LES, presentan un componente genético complejo y son producidas por la combinación de múltiples mutaciones en varios genes. Además, en este tipo de herencia poligénica se produce el fenómeno del umbral de riesgo: la enfermedad se desarrolla cuando la contribución de múltiples factores genéticos supera este umbral<sup>69</sup>. Gracias a estos dos conceptos, consideramos que los pacientes con LES presentan un número de alelos de riesgo mayor que controles y similar entre ellos sin importar la procedencia, debido a que el

hecho de padecer la enfermedad hace que tengan una mayor similitud en los genes asociados al LES que los controles.

Otro aspecto importante, es que 7 de los SNPs estudiados muestran clinas norte-sur dentro de Europa teniendo en cuenta sus frecuencias alélicas. Solamente una de ellas, la presente en el gen *PTPN22*, se había descrito en estudios previos<sup>68</sup>. Además, no solamente encontramos una clina norte-sur en estos SNPs individuales, sino que también en la distribución de la suma general de alelos de riesgo al LES tanto en casos como en controles a lo largo de Europa. Este resultado nos aporta seguridad a la hora de la determinación geográfica de nuestros subgrupos sur y centro de Europa. Además, la presencia de esta clina norte-sur en la suma de alelos de riesgo en general sugiere la existencia de fuerzas de selección que actuaron sobre diferentes poblaciones europeas a lo largo de la historia como podrían ser la resistencia a infecciones concretas<sup>70, 71</sup>.

Este estudio presenta una limitación importante que consiste en el pequeño tamaño del subgrupo de muestras del centro de Europa (383 casos con LES y 463 controles) debido, en parte, a la exclusión de las colecciones de Bélgica y República Checa por problemas de heterogeneidad en el ancestro. Sin embargo, un trabajo previo ha analizado la distribución de una variable consistente en la suma de alelos de riesgo al LES del varios *loci* consiguiendo resultados realmente similares a los nuestros<sup>72</sup>. Su estudio se realizó con una amplia población de 1919 casos con LES y 4813 controles de ascendencia euro-americana con clara influencia del norte y centro de Europa. Para la creación de la variable suma de alelos de riesgo, utilizan 22 señales de asociación al LES de la cuales compartimos un total de 10 y serían un número comparable a las 20 utilizadas en nuestro estudio. Aun no siendo estas variables totalmente idénticas, los valores finales de la variable suma de alelos de riesgo deberían ser similares ya que estamos considerando que el resultado en las diferencias de tamaños de efecto encontradas se deben a un fenómeno general en la susceptibilidad al LES influenciado por la subestructura europea y no a la variación individual de algunas señales. Además, las 22 señales mostraron resultados de asociación significativa en su colección. Concretamente, sus controles mostraron una media de suma de alelos de riesgo de  $13,1 \pm 2,8$  dato comparable con nuestros controles de media  $12,8 \pm 2,7$ . Los mismos resultados obtuvimos comparando los casos con LES de su estudio con una media de  $15,1 \pm 3,1$  y nuestro casos con LES con una media de  $14,4 \pm 2,8$ . Por lo tanto, teniendo en cuenta la comparación de estos resultados, podemos considerar nuestra colección del centro de Europa representativa de su localización geográfica a pesar de su pequeño tamaño.

Finalmente, parte de los resultados de nuestro estudio también se pueden clasificar como una replicación independiente. De las 16 señales de asociación al LES estudiadas en la segunda fase de este estudio, cuatro de ellas no mostraron resultados de asociación significativa en nuestra colección de muestras teniendo en cuenta tanto las colecciones del centro como las del sur de Europa en conjunto mediante un meta-análisis con el test Mantel-Haenszel. Todas las 16 señales se incluyeron en este estudio ya que mostraban resultados de asociación claros en varios estudios previos y un tamaño de efecto mínimo de su alelo de riesgo  $OR > 1,15$  (**referencias en la Tabla Suplementaria 1 de la Publicación 2**). Las señales de los genes *ICAI*, *LYN* y *XKR6* fueron descubiertas en uno de los GWAS para LES realizados en población europea y replicados, de manera independiente, en al menos un estudio<sup>73, 74</sup>. Por otro lado, la señal de asociación del gen *PERP* con la susceptibilidad al LES solamente se había descrito en un estudio previo<sup>75</sup>. En todos los casos, nuestro estudio estaba significativamente capacitado para detectar las posibles asociaciones de estas señales.

Como conclusión, nuestro estudio describe un claro sesgo hacia un aumento en el tamaño del efecto de múltiples señales de asociación al LES en las poblaciones del sur en comparación con las poblaciones del centro de Europa. Por lo tanto, este hallazgo contribuirá a definir la epidemiología genética del LES además de proponer nuevas líneas de trabajo para el estudio de su complejo y heterogéneo componente genético.

### **1.3. Estudio de la variabilidad genética de *IRF5* en relación al LES: Publicación 3.**

Como ya se ha referido anteriormente, entre los factores genéticos de mayor importancia en la susceptibilidad al LES se encuentra el gen *IRF5*. De esta manera, *IRF5* ha mostrado resultados de asociación significativos en cinco de los seis GWAS realizados en LES hasta la fecha no sólo en población con ascendencia europea sino que también con ascendencia asiática<sup>19-21, 23, 76</sup>. Además, *IRF5* también se ha encontrado asociado al LES en población con ascendencia afro-americana<sup>77, 78</sup>. A pesar de todos estos estudios, los detalles de esta asociación no están completamente claros. Por ello, nos planteamos estudiar la influencia de la variabilidad genética presente en *IRF5* en relación con la susceptibilidad al LES mediante un estudio de asociación caso-control en una colección de 1614 controles y 1383 casos con LES procedentes de 14 colecciones europeas (**Publicación 3**). Para cubrir la variabilidad genética de *IRF5* y secuencias vecinas genotipamos 9 *tag* SNPs. Además, en este estudio también se introdujeron todos los polimorfismos funcionales descritos en relación al gen *IRF5* hasta la fecha: SNP rs2004640, indel del exón 6 y el SNP rs10954213 (poly A). Todos los polimorfismos incluidos en este estudio, a excepción del indel del exón 6, mostraron asociación significativa en relación a la susceptibilidad al LES resultado del meta-análisis mediante el test Mantel-Haenszel. Estas

diferencias significativas entre frecuencias alélicas se extienden unas 25 Kb aproximadamente ya que es la región que cubren los *tag* SNPs seleccionados. Estas evidencias junto al bajo valor de  $r^2$  entre los *tag* SNPs seleccionados, nos dan una idea de que el gen *IRF5* contiene múltiples polimorfismos que modifican la susceptibilidad al LES, y muchos de ellos aún no conocidos.

Los resultados obtenidos en este estudio permitieron clarificar y definir en más detalle la contribución del gen *IRF5* en relación con la susceptibilidad al LES. Especialmente, nuestro estudio fue el primero en definir dos efectos independientes y opuestos dentro del gen *IRF5*. Por un lado, encontramos el efecto protector definido por el alelo menor C del SNP rs729302 en los haplotipos #1 y #2 significativamente asociados con la protección al LES. Solamente en un estudio anterior se había sugerido la existencia de un haplotipo protector independiente definido por el alelo G del SNP funcional rs2004640<sup>37</sup>. Sin embargo, en este estudio no quedaba claro si este efecto protector era independiente del de riesgo, siendo debido a un fenómeno de compensación resultante del efecto de susceptibilidad del alelo T del mismo SNP rs2004640. Concretamente, en nuestro estudio, el alelo G del SNP funcional rs2004640 está presente no sólo en haplotipos protectores #1 y #2, sino que también en el haplotipo neutro #3. Sin embargo, el efecto protector asociado al SNP rs729302 ya se había observado en varios estudios previos al nuestro, pero no se había considerado por separado, ni demostrado que era independiente<sup>34, 35, 39</sup>. Con posterioridad, se ha confirmado en poblaciones asiáticas donde este SNP rs729302 es mucho más frecuente<sup>79, 80</sup>. Además, en dos de los GWAS para LES realizados hasta la fecha, hemos encontrado el mismo efecto protector del SNP rs729302 en población caucásica<sup>20</sup> y asiática<sup>23</sup>. Por lo tanto, el efecto protector asociado al alelo raro del SNP rs729302 está sólidamente confirmado.

En relación a la funcionalidad de este *tag* SNP protector rs729302, no se había descrito ningún posible papel que explicara su relación con la protección al LES. Analizamos los datos de LD de la colección CEPH de ascendencia europea del Proyecto Internacional HapMap y no encontramos ningún polimorfismo con una alta  $r^2$  con él para poder así funcionar como su *tag* SNP. Además, este *tag* SNP protector rs729302 está localizado en la región promotora de *IRF5* 9 Kb *upstream* del inicio del gen. Esta localización en promotor le aporta un posible papel funcional distorsionando la unión de factores de transcripción y por lo tanto, promoviendo un posible cambio en la expresión del gen *IRF5*. Esta hipótesis se testará en trabajos posteriores en nuestro laboratorio.

Por otro lado, encontramos un efecto de susceptibilidad independiente definido por el alelo menor C del *tag* SNP rs10488631 en el haplotipo #6 significativamente asociado con el aumento de la susceptibilidad al LES. Este *tag* SNP de susceptibilidad rs10488631 no presenta

ninguna repercusión funcional conocida y está localizado 4 Kb *downstream* del final del gen *IRF5*. Mediante el análisis de el mapa de LD de la colección CEPH de ascendencia europea del Proyecto Internacional HapMap, pudimos observar que existen un total de 10 polimorfismos presentes en el gen *TNPO3* con un  $r^2 > 0,8$  con el SNP rs10488631, pudiendo él así funcionar como su *tag* SNP. La hipótesis que planteamos consiste en que alguno de estos 10 SNPs presentaría una repercusión funcional que explicaría la asociación del gen *IRF5* con la susceptibilidad al LES. Sin embargo, esta hipótesis pierde importancia ya que el gen *TNPO3*, implicado en el transporte de proteínas del spliceosoma al núcleo, no ha presentado evidencias previas de su papel en la patogenia del LES. Además, ninguno de los 10 SNPs presentó una posible repercusión funcional testada mediante herramientas bioinformáticas como la aplicación Pupasuite (<http://pupasuite.bioinfo.cipf.es/>).

Una mejor alternativa es la presencia de la combinación epistásica “T-in-A” de los polimorfismo funcionales de *IRF5* conocidos hasta esa fecha: SNP rs2004640, indel del exón 6 y SNP rs10954213 (poly A). Esta combinación alélica epistásica está únicamente presente en el haplotipo de riesgo #6, y por lo tanto, el SNP rs10488631 funcionaría perfectamente como su *tag* SNP. Concretamente, esta combinación epistásica “T-in-A” promueve la expresión de la isoforma de *IRF5* que contiene el exón 1B, una extensa región rica en prolina dentro del exón 6 y por último, una cola de poly A corta que correlacionada con una mayor expresión. Es importante destacar que ninguno de estos polimorfismos funcionales ejerce un papel individual en la explicación de la asociación del gen *IRF5* con la susceptibilidad al LES. Al mismo tiempo y posteriormente a la publicación de nuestro estudio, varios trabajos validaron la existencia de esta combinación epistásica entre los polimorfismos funcionales incluidos en este estudio y otros descubiertos con posterioridad que se explicarán seguidamente en la discusión. Por lo tanto, todos estos polimorfismos con repercusión funcional se han organizado en la distribución haplotípica de *IRF5* de tal forma que el haplotipo de riesgo para la enfermedad está constituido por el alelo T del SNP rs2004640, el alelo A del SNP rs10954213 y el alelo inserción del indel del exón 6. Este haplotipo se ha definido en varios estudios independientes y se ha asociado no sólo con una mayor expresión de *IRF5* sino que también de IFN $\alpha$ <sup>33-35, 37, 38, 81, 82</sup>.

Como conclusión de este trabajo, podemos decir que describimos por primera vez dos efectos independientes y opuestos dentro del *locus* de riesgo al LES, *IRF5*. Sin embargo, la repercusión de estos dos efectos no está clara. Por un lado, se podría explicar la repercusión funcional del efecto de susceptibilidad mediante el efecto epistásico “T-in-A”. Pero por otro lado, la repercusión funcional del efecto protector no presenta una diana concreta.

#### 1.4. Análisis de la regulación de la expresión génica de *IRF5* mediante datos de expresión procedentes de estudios de *microarrays*: Publicación 4.

En nuestro estudio previo (**Publicación 3**), describimos los dos efectos independientes y opuestos presentes en el gen *IRF5* en relación al LES. Sin embargo, la explicación funcional de estos dos efectos en relación a la patogenia del LES está aún por explicar. Por un lado, el efecto de susceptibilidad ha sido correlacionado con la combinación alélica epistásica “T-in-A” entre los 3 polimorfismos funcionales conocidos hasta la fecha: SNP rs2004640, indel del exón 6 y el SNP rs10954213 (**Publicación 3**). Esta combinación alélica de los tres polimorfismos funcionales marcando la susceptibilidad ya se había sugerido en estudios previos<sup>34, 35, 37</sup>. Además, a partir del descubrimiento del polimorfismo funcional CGGGG indel localizado en la región promotora del gen, se consideró que este polimorfismo determinaría el aumento de riesgo al LES en conjunto y de manera independiente al SNP de susceptibilidad rs10488631<sup>39</sup>.

Por otro lado, no existe ninguna evidencia sobre cuál es el mecanismo funcional de la protección del SNP rs729302 frente al LES. En primer lugar, no existe ningún polimorfismo funcional conocido en alta correlación con el SNP protector rs729302. Sin embargo, para los primeros autores que abordaron el estudio del gen *IRF5*, el efecto protector observado era debido a la ausencia de los alelos de susceptibilidad de los tres polimorfismos funcionales conocidos hasta la fecha (SNP rs2004640, indel del exón 6 y SNP rs10954213)<sup>33, 83</sup>. Sin embargo, como ya se ha visto en nuestro estudio previo, esta posibilidad queda descartada ya que se establece a la protección como un efecto independiente de la susceptibilidad (**Publicación 3**). La descripción del CGGGG indel le aportó una explicación funcional a la asociación del gen *IRF5* con la susceptibilidad al LES. En concreto, el alelo de riesgo inserción crea un lugar de unión adicional para el factor de transcripción SP1 junto con un aumento de la expresión de *IRF5*<sup>39</sup>. Según los autores de este trabajo, el CGGGG indel y el SNP rs10488631, que define nuestro haplotipo de susceptibilidad #6 (**Publicación 3**), serían suficientes para explicar la asociación de *IRF5* con el LES, ya que el efecto protector del SNP rs729302 vendría recogido por el alelo delección del CGGGG indel<sup>39</sup>.

En base a las evidencias previas, podemos decir que ninguno de los dos efectos independientes y opuestos descritos en *IRF5* en relación al LES, presenta una relación clara con ninguno de los polimorfismos funcionales conocidos hasta la fecha o con un cambio en la función del gen. Por lo tanto, este trabajo aporta, por primera vez el análisis directo de los datos de expresión de *IRF5* obtenidos de varios estudios realizados con *microarrays* de expresión a nivel de genoma completo<sup>40-43</sup> en relación a los cuatro polimorfismos de *IRF5* descritos hasta la fecha. Además de los 4 polimorfismos funcionales, este estudio proporciona una amplia

cobertura genética con más de 100 polimorfismos incluidos dentro del bloque de LD del gen *IRF5*. Es importante destacar que dentro de esos 100 polimorfismos, en nuestro estudio también incluimos los genotipos de alrededor de 30 polimorfismos localizados en la región promotora de *IRF5* para así poder encontrar algún polimorfismo funcional del que pueda ser *tag* SNP el rs729302 cuyo alelo raro define el efecto protector previamente descrito (**Publicación 3**).

La utilización de datos de expresión de múltiples estudios les aporta reproducibilidad y generalidad a nuestros resultados. Así la utilización de múltiples sondas que hibridan con diferentes zonas del gen *IRF5*, nos permitió seleccionar los datos de expresión más representativos del gen. Sin embargo, las diferencias presentes entre experimentos dieron lugar a diferencias en nuestros resultados. Por ejemplo, la proporción de la variabilidad de los datos de expresión explicada por cada uno de los mejores modelos de regresión lineal presenta un amplio rango de variación (0,28-0,80). Estas diferencias tienen difícil explicación. Podemos asegurar que no se deben a diferencias en el tamaño muestral, a diferencias en el escalado de los datos de expresión por estudio o a su dispersión. Sin embargo, al comparar los mejores modelos de regresión obtenidos con cada una de las 8 sondas del estudio de Kwan T. et al<sup>42</sup>, obtuvimos un rango similar de diferencias. Ya que la hibridación de estas sondas se realizó en el mismo experimento, excluimos cualquier interferencia a nivel experimental. Por lo tanto, consideramos que las diferencias proceden de la naturaleza de las sondas. Por lo tanto, una posible explicación se basa en alguna de las limitaciones que tienen los estudios de *microarrays* de expresión: el alto ruido de fondo de la señal y el punto de saturación, ambos dependientes tanto de la región genómica a hibridar como de la longitud de la sonda<sup>46</sup>.

A pesar de estas diferencias, todos los mejores modelos de regresión lineal para cada set de datos de expresión analizado, comparten parcialmente los mismos polimorfismos. De hecho, es casi imposible diferenciar entre los mejores modelos de regresión lineal analizados de manera intercambiada en cada set de datos (**Figura 3 de la Publicación 4**). Sin embargo, hay una combinación de polimorfismos que es claramente superior a las otras en todas estas comparaciones. Esta combinación consiste en el SNP rs10954213, SNP rs3807306 y el SNP rs17424179. Resumiendo, este modelo es el mejor descrito para los datos de expresión S<sup>43</sup>, no significativamente peor del mejor modelo para los datos de expresión de K8<sup>42</sup>, el segundo mejor modelo para los datos de D<sup>41</sup> y el tercer mejor modelo para los datos de C<sup>40</sup>. Por último, es importante destacar que también consiste en el mejor modelo para la sonda consenso KSC la cual nos daría una idea generalizada de la expresión de *IRF5* en la colección CEPH de LCLs del Proyecto Internacional HapMap (**Figura 3 de la Publicación 4**).

En relación al SNP rs17424179, no hay ninguna evidencia funcional ni de asociación previa que pueda describir su influencia en la regulación de la expresión de *IRF5*. Sin embargo, el SNP rs3807306, localizado en el intrón 1 del gen, se ha encontrado asociado al LES tanto en población caucásica<sup>39</sup> como en población africana<sup>77</sup>. Concretamente, en uno de estos estudios se ha considerado que el posible efecto del SNP rs3807306 es debido al CGGGG indel funcionando como su *proxy* ya que presenta un alto valor de  $r^2$  con éste ( $r^2 = 0,75$  en nuestro estudio)<sup>39</sup>. Sin embargo, esta hipótesis pierde validez en nuestro estudio ya que ambos polimorfismos están incluidos en nuestros análisis. Además, este SNP rs3807306 no presenta unión alélica diferencial a factores de transcripción testado mediante EMSA con extractos nucleares de sangre total de controles sanos<sup>44</sup>. Por lo tanto, este SNP rs3807306 presenta un papel funcional en la regulación de la expresión de *IRF5* mediante algún mecanismo sin involucrar una diferencia en la unión de factores de transcripción.

El polimorfismo con un papel claramente dominante en la explicación de la variación de los datos de expresión analizados, consistió en el SNP rs10954213 (poly A). Esta dominancia se observa no sólo en su comparación individual con los polimorfismos funcionales SNP rs2004640 y el CGGGG indel, sino que también en las comparaciones dos a dos (**Figura 4 de la Publicación 4**). Por lo tanto, parece que la contribución al modelo de los demás polimorfismos funcionales es redundante a la del SNP rs10954213. Sin embargo, este fenómeno no se cumple en el set de datos de D<sup>41</sup> ya que los modelos rs2004640-rs10954213 y CGGGG indel-rs10954213 no fueron significativamente diferentes al mejor modelo de regresión lineal descrito con los datos de D<sup>41</sup>. Por lo tanto, no sabemos cuál de los sets de muestras, LCLs de la colección CEPH del Proyecto Internacional HapMap<sup>40, 42, 43</sup> o las LCLs de la colección del estudio de asma<sup>41</sup>, es más representativa de la población europea. De todos modos, como ya se ha comentado en la introducción de este trabajo, muchas evidencias previas validan el papel de SNP rs10954213 (poly A) en la regulación de la expresión de *IRF5*<sup>33, 37, 38</sup>. Sin embargo, existe algún estudio en el que este efecto no se ha replicado significativamente por falta de poder estadístico debido al pequeño tamaño de sus comparaciones<sup>81</sup>. Sin embargo, en todos los análisis de este estudio se aprecia una tendencia hacia el aumento de la expresión de *IRF5* con el alelo de riesgo A del SNP rs10954213 y un aumento significativo en la producción proteica de *IRF5* en pacientes con LES<sup>81</sup>.

Por último, nuestros resultados confirman la asociación del aumento de expresión de *IRF5* en relación al haplotipo de susceptibilidad #6 en los cuatro set de datos analizados (**Figura 5 de la Publicación 4**). Este efecto está ampliamente demostrado tanto en pacientes con LES como en controles<sup>38, 81</sup>. Además, es la base de la explicación funcional de *IRF5* en el patogenicidad del LES mediante la cual el haplotipo de riesgo aumentaría la expresión del gen aumentando por consiguiente la expresión de la ruta del IFN de tipo I<sup>82, 84</sup>.



Sin embargo, nuestros estudios dejan la puerta abierta tanto al descubrimiento de nuevos polimorfismos funcionales como a diferentes modificaciones que influyen la regulación de la expresión de *IRF5*. Esto es debido a la falta de la completa correlación entre los dos efectos protección, marcado por los haplotipos #1 y #2, y susceptibilidad, marcada por el haplotipo #6, con la expresión de *IRF5* (**Figura 5 de la Publicación 4**). Además, tampoco encontramos ninguna correlación entre la expresión descrita por los mejores modelos de regresión en cada set de datos con ninguno de los haplotipos, tanto de susceptibilidad como de protección (**Tabla suplementaria 3 de la Publicación 4**). Por lo tanto, podemos decir que la regulación de la expresión de *IRF5* por medio de los polimorfismos funcionales conocidos hasta la fecha, no es suficiente para la explicación total de su asociación con el LES aún encontrando una consistente asociación del haplotipo de riesgo #6 con el aumento de su expresión.

### 1.5. Búsqueda de polimorfismos funcionales en la región promotora de *IRF5*.

Como ya se ha referido en apartados previos, no existe ninguna evidencia clara sobre cuál es el mecanismo funcional que explique la protección del gen *IRF5* frente al LES. Por lo tanto, el paso inicial de nuestra búsqueda consistió en evaluar funcionalmente el SNP rs729302 cuyo alelo raro C caracteriza a los haplotipos de protección #1 y #2 previamente descritos (**Publicación 3**). Este SNP rs729302 está localizado en la región promotora de *IRF5*, 9 Kb *upstream* del inicio del gen. Por lo tanto, asumimos un posible efecto regulador de la expresión génica mediante la unión de factores de transcripción. Sin embargo, los ensayos EMSA realizados no mostraron unión alélica diferencial a factores de transcripción (**Figura 1**). Por lo tanto, debido a estos resultados negativos nos planteamos la hipótesis de que el SNP rs729302 funcionaría como un *tag* SNP de otro polimorfismo desconocido con efecto funcional.

Teniendo como objetivo la hipótesis previa, decidimos estudiar en detalle la región promotora de *IRF5* debido a la localización del SNP protector rs729302 en esa misma región. Como método de búsqueda utilizamos los datos de LD de estos nuevos polimorfismos con los *tag* SNPs de *IRF5* previamente estudiados (**Publicación 3**). De esta forma establecimos un puente entre los nuevos polimorfismos y la distribución de asociación genética previamente detallada (**Publicación 3**). A partir de estos datos de LD establecimos varios criterios de selección de polimorfismos posiblemente funcionales en base a su relación con el efecto protector previamente descrito (**Publicación 3**). El primer criterio de selección consistió en seleccionar todos los polimorfismos con alta correlación con el SNP protector rs729302, ya que son los que de forma más clara podrían explicar su efecto. Por lo tanto, se testaron mediante ensayos EMSA un total de 8 SNPs, todos ellos con una  $r^2 > 0,7$  con el SNP protector rs729302

(Figura 3) Seguidamente, en un segundo grupo, se seleccionaron polimorfismos que sin muy buena correlación con el SNP rs729302 agrupaban un número considerable de los subhaplotipos en los que la inclusión de nuevos polimorfismos había dividido a los dos haplotipos protectores #1 y #2 (Tabla 2). Teniendo en cuenta este criterio testamos funcionalmente mediante ensayos EMSA otros 9 SNPs adicionales presentes en la región promotora de *IRF5*.

Del total de 17 SNPs seleccionados mediante estos dos criterios, solamente 4 de ellos mostraron unión alélica diferencial a factores de transcripción: rs12706860, rs13245639, rs3778754 y el rs3807307. Los 13 restantes fueron claramente negativos (Figuras suplementarias 1-13).

De todos modos, es importante tener en cuenta cuáles son las limitaciones de los ensayos EMSA para predecir la repercusión funcional de un polimorfismo que ha mostrado unión alélica diferencial en esta búsqueda funcional inicial. Los ensayos EMSA son en sí insuficientes para determinar si un polimorfismo tiene repercusión funcional o no. Sólo nos indican que los polimorfismos que presentan cambios alelo-específicos en la unión de factores de transcripción podrían llegar a ser funcionales. De hecho, se necesitan otro tipo de ensayos para establecer si esta posibilidad es cierta o no. Por lo tanto, el paso siguiente en nuestra búsqueda consistió en analizar mediante ensayos gen *reporter* de luciferasa cómo se afecta la expresión de este gen luciferasa por los alelos de los SNPs con unión diferencial en los ensayos EMSA. Sólo uno de los 4 SNPs analizados mostró diferencias alelo-específicas en la regulación de la expresión de la luciferasa de manera significativa: rs13245639 (Figura 6). Podemos decir que esta es una prueba más convincente, pues ya implica un sistema *in vivo*, aunque sólo celular, y todos los elementos necesarios para la transcripción. Por el contrario, el ensayo EMSA es un sistema puramente *in vitro* que únicamente nos informa sobre la interacción molecular.

Otro resultado importante dentro de este apartado es la confirmación adicional de la funcionalidad del polimorfismo CGGGG indel. Como ya se ha referido anteriormente, este CGGGG indel mostró mediante ensayos EMSA que la inserción 4xCGGGG crea un lugar de unión adicional en la región promotora de *IRF5* para el factor de transcripción SP1 (*specificity protein 1*)<sup>30, 39, 44</sup>. Además, ensayos *reporter* con mini-genes y análisis *immunoblot* muestran que las células homocigotas para la inserción 4xCGGGG presentan niveles de expresión de *IRF5* que contiene el exón 1A entre 2 y 10 veces superiores a las células homocigotas para la delección 3xCGGGG<sup>39</sup>. Del mismo modo que con los demás polimorfismos seleccionados, testamos funcionalmente este CGGGG indel mediante ensayos EMSA y ensayos gen *reporter* de luciferasa. De manera que confirmamos la repercusión funcional del CGGGG indel

previamente descrita no sólo en relación a su unión alélica diferencial a factores de transcripción, sino que también su papel en la regulación de la expresión génica de *IRF5*.

Del mismo modo que analizamos los datos de expresión de *IRF5* obtenidos a partir de estudio con *microarrays* en nuestro trabajo previo (**Publicación 4**), decidimos analizar otros datos de expresión a partir de un estudio de secuenciado del transcriptoma de LCLs de la colección CEPH del Proyecto Internacional HapMap<sup>45</sup>. El secuenciado del transcriptoma se basa en secuenciación de alto rendimiento denominada RNA-seq. Es importante destacar que esta tecnología ofrece una serie de ventajas con respecto a los estudios con *microarrays* que le aportan una alta precisión a la hora de evaluar el nivel de expresión. La primera de estas ventajas es que la técnica RNA-seq no está sujeta al conocimiento previo de los transcritos analizados. En el caso de los *microarrays* de expresión, es necesario el conocimiento de la secuencia de los transcritos a estudiar para poder diseñar las sondas de hibridación. Otra ventaja importante que le aporta exactitud a los datos de RNA-seq consiste en que esta técnica no presenta ruido de fondo ya que en el proceso de alineamiento de los *reads*, éstos lo hacen únicamente con una región específica del genoma. De manera contraria, los *microarrays* de expresión generan mucho ruido de fondo debido a la unión inespecífica de las sondas lo que distorsiona sus resultados. Por último, una de las mayores ventajas que presenta la técnica RNA-seq consiste en que nos proporciona resultados digitales de expresión (número de *reads* que hibridan con una determinada región genómica), lo que amplía significativamente su rango dinámico de detección. Con lo cual podemos detectar genes que se expresan tanto en altos como en bajos niveles. Sin embargo, el rango dinámico de detección de los *microarrays* de expresión es mucho más limitado. Por un lado, los genes que se expresan en bajos niveles no son detectados a causa del ruido de fondo, y por el otro, los genes que se expresan en alto niveles saturarán la señal de la sonda<sup>46</sup>.

Al igual que en nuestro trabajo previo con *microarrays* (**Publicación 4**), a partir de la obtención de estos resultados de expresión de *IRF5* de alta precisión con RNA-seq, analizamos cuál sería el mejor modelo de regresión lineal con polimorfismos de *IRF5* que explique la mayor varianza posible del fenotipo. La gran diferencia con el estudio de *microarrays*, es que en este caso utilizamos la base de datos de genotipado *1000 Genomes*. El genotipado de *1000 Genomes* también se basa en secuenciación de alto rendimiento, lo que permite no sólo un genotipado preciso de los polimorfismos, sino que también el descubrimiento de nuevos polimorfismos previamente no descritos y por lo tanto no incluidos en los *arrays* o *chips* de genotipado a nivel de genoma completo. Por lo tanto, en este análisis duplicamos el número de SNPs incluidos en comparación con el Proyecto Internacional HapMap, y además, incluimos 29

indeles localizados dentro del bloque de LD de *IRF5* que no estaban cubiertos con los *chips* de genotipado a nivel de genoma completo. Como resultado de este análisis, obtuvimos un modelo de regresión lineal que explica una pequeña fracción de la varianza de estos datos de expresión en comparación con los modelos obtenidos en el estudio con *microarrays* (**Publicación 4**). Sin embargo, la mayor contribución al modelo la ejerce el indel rs11269962 localizado en la región promotora de *IRF5* previamente no descrito en ningún estudio. Por lo tanto, debido a su localización en el promotor de *IRF5* y a su naturaleza como indel de 14 pb probablemente con una mayor repercusión funcional que un SNP, decidimos testar funcionalmente al indel rs11269962 mediante ensayos EMSA y gen *reporter* de luciferasa. En base a los resultados del ensayo EMSA y ensayo gen *reporter* de luciferasa, hipotetizamos que la unión específica que observamos con el alelo inserción del indel rs11269962 promueve una inhibición del expresión del gen de la luciferasa. Por último, es importante destacar que, del mismo modo que en los ensayos gen *reporter* de luciferasa, el alelo delección de este indel rs11269962 promueve, en el modelo de regresión lineal, un aumento de la expresión del fenotipo final de *IRF5* obtenido mediante RNA-seq.

Finalmente, la búsqueda de polimorfismos funcionales mediante el estudio del mapa de LD de *IRF5*, resultó en la validación funcional de 3 polimorfismos funcionales: rs13245639, CGGGG indel y el indel rs11269962. Dos de ellos, rs13245639 y el indel rs11269962 no habían sido descritos en ningún estudio previo. Sin embargo, los resultados del CGGGG indel consistieron en una replicación independiente de su papel funcional previamente descrito<sup>30, 39, 44</sup>. Todos estos polimorfismos están localizados en la región promotora del gen pudiendo así explicar funcionalmente el efecto protector de *IRF5* frente al LES (**Publicación 3**). Por lo tanto, estos polimorfismos serán incluidos en el estudio final de asociación caso-control de *IRF5*. De esta manera analizaremos su asociación y su distribución haplotípica en relación con la enfermedad.

**CONCLUSIONES.**

---

1. Confirmamos la mayor prevalencia de fotosensibilidad y la menor prevalencia del desorden inmunológico en los pacientes con LES de ascendencia centro-europea en relación con los de ascendencia sur-europea. Además, observamos que la prevalencia de úlceras orales está especialmente aumentada en pacientes de LES con ascendencia española o portuguesa. Estos resultados corroboran la influencia de la subestructura genética europea en la heterogeneidad clínica del LES.
2. Nuestros resultados muestran la existencia de un claro sesgo hacia un aumento en el tamaño del efecto de múltiples señales de asociación al LES en las poblaciones del sur en comparación con las poblaciones del centro de Europa. Además, también se describe la presencia de clinas de la distribución de los alelos de riesgo para 7 de los SNPs estudiados: rs3131379 (*MSH5*), rs2187668 (*HLA-DQA1*), rs2476601 (*PTPN22*), rs6445975 (*PXK*), rs1143679 (*ITGAM*), rs4963128 (*KIAA1542*) y rs17435 (*MECP2*).
3. Confirmamos la asociación del gen *IRF5* al LES en una colección de muestras europeas independiente. Esta asociación presenta dos efectos independientes y opuestos. Por un lado, la protección viene definida por el SNP rs729302 localizado en la región promotora del gen. Por otro lado, la susceptibilidad viene definida por el SNP rs10488631 el cual funciona como *tag* SNP de la combinación alélica epistásica “T-in-A” de los polimorfismos funcionales: rs2004640, indel del exón 6 y rs10954213.
4. Mediante el análisis de la expresión de *IRF5* obtenida a partir de 4 estudios realizados con *microarrays*, encontramos que parte de su variabilidad puede ser explicada por un pequeño grupo de polimorfismos en el que presenta una clara dominancia el SNP funcional rs10954213 (poly A). Sin embargo, la regulación de la expresión de *IRF5* por medio de los polimorfismos funcionales conocidos hasta la fecha, no es suficiente para la explicación total de su asociación con el LES. Por último, encontramos una consistente asociación significativa del haplotipo de riesgo #6 con el aumento de la expresión de *IRF5* en los 4 set de datos de expresión analizados.
5. Mediante el estudio del mapa de LD de *IRF5* y su región promotora, encontramos dos nuevos polimorfismos con repercusión funcional validada mediante ensayos EMSA y gen *reporter* de luciferasa localizados en la región promotora del gen: SNP rs13245639 y el indel rs11269962. Además, replicamos y confirmamos la repercusión funcional del CGGGG indel previamente estudiado.

## **BIBLIOGRAFÍA.**

---

## Bibliografia

1. Simard,J.F. & Costenbader,K.H. What can epidemiology tell us about systemic lupus erythematosus? *Int. J. Clin. Pract.* **61**, 1170-1180 (2007).
2. Wakeland,E.K., Liu,K., Graham,R.R., & Behrens,T.W. Delineating the genetic basis of systemic lupus erythematosus. *Immunity.* **15**, 397-408 (2001).
3. Johnson,A.E., Gordon,C., Palmer,R.G., & Bacon,P.A. The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis Rheum.* **38**, 551-558 (1995).
4. Petri,M. Epidemiology of systemic lupus erythematosus. *Best. Pract. Res. Clin. Rheumatol.* **16**, 847-858 (2002).
5. Hochberg,M.C. Prevalence of systemic lupus erythematosus in England and Wales, 1981-2. *Ann. Rheum. Dis.* **46**, 664-666 (1987).
6. Hopkinson,N.D., Doherty,M., & Powell,R.J. Clinical features and race-specific incidence/prevalence rates of systemic lupus erythematosus in a geographically complete cohort of patients. *Ann. Rheum. Dis.* **53**, 675-680 (1994).
7. Nived,O., Sturfelt,G., & Wollheim,F. Systemic lupus erythematosus in an adult population in southern Sweden: incidence, prevalence and validity of ARA revised classification criteria. *Br. J. Rheumatol.* **24**, 147-154 (1985).
8. Feletar,M., Ibanez,D., Urowitz,M.B., & Gladman,D.D. The impact of the 1997 update of the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus: what has been changed? *Arthritis Rheum.* **48**, 2067-2069 (2003).
9. Hochberg,M.C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1725 (1997).
10. Tan,E.M. *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271-1277 (1982).
11. Sherer,Y., Gorstein,A., Fritzler,M.J., & Shoenfeld,Y. Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients. *Semin. Arthritis Rheum.* **34**, 501-537 (2004).
12. Teubner,A. *et al.* [Prevalence of circulating autoantibodies in healthy individuals]. *Med. Klin. (Munich)* **97**, 645-649 (2002).
13. Arbuckle,M.R., Reichlin,M., Harley,J.B., & James,J.A. Shared early autoantibody recognition events in the development of anti-Sm B/B' in human lupus. *Scand. J. Immunol.* **50**, 447-455 (1999).
14. Arbuckle,M.R. *et al.* Development of anti-dsDNA autoantibodies prior to clinical diagnosis of systemic lupus erythematosus. *Scand. J. Immunol.* **54**, 211-219 (2001).



15. Arbuckle,M.R. *et al.* Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N. Engl. J. Med.* **349**, 1526-1533 (2003).
16. Boh,E.E. Neonatal lupus erythematosus. *Clin. Dermatol.* **22**, 125-128 (2004).
17. Hooks,J.J., Moutsopoulos,H.M., & Notkins,A.L. Circulating interferon in human autoimmune diseases. *Tex. Rep. Biol. Med.* **41**, 164-168 (1981).
18. Niewold,T.B., Rivera,T.L., Buyon,J.P., & Crow,M.K. Serum type I interferon activity is dependent on maternal diagnosis in anti-SSA/Ro-positive mothers of children with neonatal lupus. *Arthritis Rheum.* **58**, 541-546 (2008).
19. Ben-Ami,S.D., Blank,M., & Altman,A. [The clinical importance of anti-ribosomal-P antibodies]. *Harefuah* **149**, 794-7, 810 (2010).
20. Wilson,W.A. *et al.* International consensus statement on preliminary classification criteria for definite antiphospholipid syndrome: report of an international workshop. *Arthritis Rheum.* **42**, 1309-1311 (1999).
21. Hooks,J.J. *et al.* Immune interferon in the circulation of patients with autoimmune disease. *N. Engl. J. Med.* **301**, 5-8 (1979).
22. Baechler,E.C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. U. S. A* **100**, 2610-2615 (2003).
23. Bennett,L. *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J. Exp. Med.* **197**, 711-723 (2003).
24. Preble,O.T., Black,R.J., Friedman,R.M., Klippel,J.H., & Vilcek,J. Systemic lupus erythematosus: presence in human serum of an unusual acid-labile leukocyte interferon. *Science* **216**, 429-431 (1982).
25. Niewold,T.B., Hua,J., Lehman,T.J., Harley,J.B., & Crow,M.K. High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. *Genes Immun.* **8**, 492-502 (2007).
26. Banchereau,J. & Pascual,V. Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity.* **25**, 383-392 (2006).
27. Palucka,A.K., Banchereau,J., Blanco,P., & Pascual,V. The interplay of dendritic cell subsets in systemic lupus erythematosus. *Immunol. Cell Biol.* **80**, 484-488 (2002).
28. Merrill,J.T. *et al.* Safety profile and clinical activity of sifalimumab, a fully human anti-interferon alpha monoclonal antibody, in systemic lupus erythematosus: a phase I, multicentre, double-blind randomised study. *Ann. Rheum. Dis.* **70**, 1905-1913 (2011).
29. Obermoser,G. & Pascual,V. The interferon-alpha signature of systemic lupus erythematosus. *Lupus* **19**, 1012-1019 (2010).
30. Pascual,V., Farkas,L., & Banchereau,J. Systemic lupus erythematosus: all roads lead to type I interferons. *Curr. Opin. Immunol.* **18**, 676-682 (2006).

31. Ronnblom,L., Alm,G.V., & Eloranta,M.L. Type I interferon and lupus. *Curr. Opin. Rheumatol.* **21**, 471-477 (2009).
32. Yao,Y. *et al.* Neutralization of interferon-alpha/beta-inducible genes and downstream effect in a phase I trial of an anti-interferon-alpha monoclonal antibody in systemic lupus erythematosus. *Arthritis Rheum.* **60**, 1785-1796 (2009).
33. Vilcek,J. Fifty years of interferon research: aiming at a moving target. *Immunity.* **25**, 343-348 (2006).
34. Seitz,H.M. & Matsushima,G.K. Dendritic cells in systemic lupus erythematosus. *Int. Rev. Immunol.* **29**, 184-209 (2010).
35. Fitzgerald-Bocarsly,P., Dai,J., & Singh,S. Plasmacytoid dendritic cells and type I IFN: 50 years of convergent history. *Cytokine Growth Factor Rev.* **19**, 3-19 (2008).
36. Heil,F. *et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526-1529 (2004).
37. Hemmi,H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740-745 (2000).
38. Ronnblom,L. The type I interferon system in the etiopathogenesis of autoimmune diseases. *Ups. J. Med. Sci.* **116**, 227-237 (2011).
39. Ronnblom,L. The type I interferon system in the etiopathogenesis of autoimmune diseases. *Ups. J. Med. Sci.* **116**, 227-237 (2011).
40. Bave,U. *et al.* Fc gamma RIIa is expressed on natural IFN-alpha-producing cells (plasmacytoid dendritic cells) and is required for the IFN-alpha production induced by apoptotic cells combined with lupus IgG. *J. Immunol.* **171**, 3296-3302 (2003).
41. Jego,G. *et al.* Plasmacytoid dendritic cells induce plasma cell differentiation through type I interferon and interleukin 6. *Immunity.* **19**, 225-234 (2003).
42. Filaci,G. *et al.* Impairment of CD8+ T suppressor cell function in patients with active systemic lupus erythematosus. *J. Immunol.* **166**, 6452-6457 (2001).
43. Oishi,Y. *et al.* Selective reduction and recovery of invariant Valpha24JalphaQ T cell receptor T cells in correlation with disease activity in patients with systemic lupus erythematosus. *J. Rheumatol.* **28**, 275-283 (2001).
44. Singh,R.R. *et al.* Induction of autoantibody production is limited in nonautoimmune mice. *J. Immunol.* **169**, 587-594 (2002).
45. Yan,B. *et al.* Dysfunctional CD4+,CD25+ regulatory T cells in untreated active systemic lupus erythematosus secondary to interferon-alpha-producing antigen-presenting cells. *Arthritis Rheum.* **58**, 801-812 (2008).
46. Yang,J.Q. *et al.* Repeated alpha-galactosylceramide administration results in expansion of NK T cells and alleviates inflammatory dermatitis in MRL-lpr/lpr mice. *J. Immunol.* **171**, 4439-4446 (2003).

47. Alarcon-Segovia,D. *et al.* Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* **52**, 1138-1147 (2005).
48. Deapen,D. *et al.* A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* **35**, 311-318 (1992).
49. Deng,Y. & Tsao,B.P. Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nat. Rev. Rheumatol.* **6**, 683-692 (2010).
50. Moser,K.L., Kelly,J.A., Lessard,C.J., & Harley,J.B. Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun.* **10**, 373-379 (2009).
51. Tsao,B.P. The genetics of human systemic lupus erythematosus. *Trends Immunol.* **24**, 595-602 (2003).
52. Wandstrat,A. & Wakeland,E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat. Immunol.* **2**, 802-809 (2001).
53. Wong,M. & Tsao,B.P. Current topics in human SLE genetics. *Springer Semin. Immunopathol.* **28**, 97-107 (2006).
54. Ballestar,E., Esteller,M., & Richardson,B.C. The epigenetic face of systemic lupus erythematosus. *J. Immunol.* **176**, 7143-7147 (2006).
55. Ballestar,E. Epigenetic alterations in autoimmune rheumatic diseases. *Nat. Rev. Rheumatol.* **7**, 263-271 (2011).
56. Zhao,S., Long,H., & Lu,Q. Epigenetic perspectives in systemic lupus erythematosus: pathogenesis, biomarkers, and therapeutic potentials. *Clin. Rev. Allergy Immunol.* **39**, 3-9 (2010).
57. Cantor,R.M. *et al.* Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23. *Arthritis Rheum.* **50**, 3203-3210 (2004).
58. Gaffney,P.M. *et al.* A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc. Natl. Acad. Sci. U. S. A* **95**, 14875-14879 (1998).
59. Gaffney,P.M. *et al.* Genome screening in human systemic lupus erythematosus: results from a second Minnesota cohort and combined analyses of 187 sib-pair families. *Am. J. Hum. Genet.* **66**, 547-556 (2000).
60. Gray-McGuire,C. *et al.* Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. *Am. J. Hum. Genet.* **67**, 1460-1469 (2000).
61. Johansson,C.M. *et al.* Chromosome 17p12-q11 harbors susceptibility loci for systemic lupus erythematosus. *Hum. Genet.* **115**, 230-238 (2004).
62. Koskenmies,S. *et al.* Linkage mapping of systemic lupus erythematosus (SLE) in Finnish families multiply affected by SLE. *J. Med. Genet.* **41**, e2-e5 (2004).

63. Lindqvist,A.K. *et al.* A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. *J. Autoimmun.* **14**, 169-178 (2000).
64. Moser,K.L. *et al.* Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. *Proc. Natl. Acad. Sci. U. S. A* **95**, 14869-14874 (1998).
65. Nath,S.K. *et al.* Linkage at 12q24 with systemic lupus erythematosus (SLE) is established and confirmed in Hispanic and European American families. *Am. J. Hum. Genet.* **74**, 73-82 (2004).
66. Olson,J.M. *et al.* A genome screen of systemic lupus erythematosus using affected-relative-pair linkage analysis with covariates demonstrates genetic heterogeneity. *Genes Immun.* **3 Suppl 1**, S5-S12 (2002).
67. Rao,S. *et al.* Linkage analysis of human systemic lupus erythematosus-related traits: a principal component approach. *Arthritis Rheum.* **44**, 2807-2818 (2001).
68. Shai,R. *et al.* Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. *Hum. Mol. Genet.* **8**, 639-644 (1999).
69. Forabosco,P. *et al.* Meta-analysis of genome-wide linkage studies of systemic lupus erythematosus. *Genes Immun.* **7**, 609-614 (2006).
70. Lee,Y.H. & Nath,S.K. Systemic lupus erythematosus susceptibility loci defined by genome scan meta-analysis. *Hum. Genet.* **118**, 434-443 (2005).
71. Graham,R.R. *et al.* Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus. *Am. J. Hum. Genet.* **71**, 543-553 (2002).
72. Tsao,B.P. *et al.* PARP alleles within the linked chromosomal region are associated with systemic lupus erythematosus. *J. Clin. Invest* **103**, 1135-1140 (1999).
73. Remmers,E.F. *et al.* STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977-986 (2007).
74. Prokunina,L. *et al.* A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat. Genet.* **32**, 666-669 (2002).
75. Ferreira-Vidal,I. *et al.* Association of PDCD1 with susceptibility to systemic lupus erythematosus: evidence of population-specific effects. *Arthritis Rheum.* **50**, 2590-2597 (2004).
76. Suarez-Gestal,M., Ferreira-Vidal,I., Ortiz,J.A., Gomez-Reino,J.J., & Gonzalez,A. Analysis of the functional relevance of a putative regulatory SNP of PDCD1, PD1.3, associated with systemic lupus erythematosus. *Genes Immun.* **9**, 309-315 (2008).
77. Begovich,A.B. *et al.* A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.* **75**, 330-337 (2004).

78. Criswell,L.A. *et al.* Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am. J. Hum. Genet.* **76**, 561-571 (2005).
79. Kyogoku,C. *et al.* Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am. J. Hum. Genet.* **75**, 504-507 (2004).
80. Smyth,D. *et al.* Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes* **53**, 3020-3023 (2004).
81. Sigurdsson,S. *et al.* Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.* **76**, 528-537 (2005).
82. Nath,S.K. *et al.* A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 152-154 (2008).
83. Frazer,K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
84. Saxena,R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-1336 (2007).
85. Dudbridge,F. & Gusnanto,A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227-234 (2008).
86. Hoggart,C.J., Clark,T.G., De,I.M., Whittaker,J.C., & Balding,D.J. Genome-wide significance for dense SNP and resequencing data. *Genet. Epidemiol.* **32**, 179-185 (2008).
87. McCarthy,M.I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9**, 356-369 (2008).
88. Tiwari,H.K. *et al.* Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. *Hum. Hered.* **66**, 67-86 (2008).
89. Reich,D., Price,A.L., & Patterson,N. Principal component analysis of genetic data. *Nat. Genet.* **40**, 491-492 (2008).
90. Price,A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904-909 (2006).
91. Birney,E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816 (2007).
92. Raney,B.J. *et al.* ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res.* **39**, D871-D875 (2011).
93. Glessner,J.T. *et al.* A genome-wide study reveals copy number variants exclusive to childhood obesity cases. *Am. J. Hum. Genet.* **87**, 661-666 (2010).

94. Grayson,B.L. *et al.* Genome-wide analysis of copy number variation in type 1 diabetes. *PLoS. One.* **5**, e15393 (2010).
95. Lee,K.W., Woon,P.S., Teo,Y.Y., & Sim,K. Genome wide association studies (GWAS) and copy number variation (CNV) studies of the major psychoses: what have we learnt? *Neurosci. Biobehav. Rev.* **36**, 556-571 (2012).
96. Meyer,K.J. *et al.* Genome-wide analysis of copy number variants in age-related macular degeneration. *Hum. Genet.* **129**, 91-100 (2011).
97. Graham,R.R. *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 1059-1061 (2008).
98. Harley,J.B. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.* **40**, 204-210 (2008).
99. Hom,G. *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* **358**, 900-909 (2008).
100. Kozyrev,S.V. *et al.* Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 211-216 (2008).
101. Han,J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234-1237 (2009).
102. Yang,W. *et al.* Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS. Genet.* **6**, e1000841 (2010).
103. Plenge,R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477-1482 (2007).
104. Thomson,W. *et al.* Rheumatoid arthritis association at 6q23. *Nat. Genet.* **39**, 1431-1433 (2007).
105. Gateva,V. *et al.* A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1228-1233 (2009).
106. Han,J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234-1237 (2009).
107. Manderson,A.P., Botto,M., & Walport,M.J. The role of complement in the development of systemic lupus erythematosus. *Annu. Rev. Immunol.* **22**, 431-456 (2004).
108. Willcocks,L.C. *et al.* Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* **205**, 1573-1582 (2008).
109. Manolio,T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009).

- 
110. Klein,J. & Sato,A. The HLA system. Second of two parts. *N. Engl. J. Med.* **343**, 782-786 (2000).
  111. Klein,J. & Sato,A. The HLA system. First of two parts. *N. Engl. J. Med.* **343**, 702-709 (2000).
  112. Graham,R.R. *et al.* Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *Eur. J. Hum. Genet.* **15**, 823-830 (2007).
  113. Sebastiani,G.D. & Galeazzi,M. Immunogenetic studies on systemic lupus erythematosus. *Lupus* **18**, 878-883 (2009).
  114. Graham,R.R., Hom,G., Ortmann,W., & Behrens,T.W. Review of recent genome-wide association scans in lupus. *J. Intern. Med.* **265**, 680-688 (2009).
  115. Fernando,M.M. *et al.* Identification of two independent risk factors for lupus within the MHC in United Kingdom families. *PLoS. Genet.* **3**, e192 (2007).
  116. Sanchez,E. *et al.* Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. *Arthritis Rheum.* **63**, 3493-3501 (2011).
  117. Barcellos,L.F. *et al.* High-density SNP screening of the major histocompatibility complex in systemic lupus erythematosus demonstrates strong evidence for independent susceptibility regions. *PLoS. Genet.* **5**, e1000696 (2009).
  118. Botto,M. & Walport,M.J. C1q, autoimmunity and apoptosis. *Immunobiology* **205**, 395-406 (2002).
  119. Fielder,A.H. *et al.* Family study of the major histocompatibility complex in patients with systemic lupus erythematosus: importance of null alleles of C4A and C4B in determining disease susceptibility. *Br. Med. J. (Clin. Res. Ed)* **286**, 425-428 (1983).
  120. Reveille,J.D., Arnett,F.C., Wilson,R.W., Bias,W.B., & McLean,R.H. Null alleles of the fourth component of complement and HLA haplotypes in familial systemic lupus erythematosus. *Immunogenetics* **21**, 299-311 (1985).
  121. Sullivan,K.E., Petri,M.A., Schmeckpeper,B.J., McLean,R.H., & Winkelstein,J.A. Prevalence of a mutation causing C2 deficiency in systemic lupus erythematosus. *J. Rheumatol.* **21**, 1128-1133 (1994).
  122. Kelly,J.A., Moser,K.L., & Harley,J.B. The genetics of systemic lupus erythematosus: putting the pieces together. *Genes Immun.* **3 Suppl 1**, S71-S85 (2002).
  123. Blanchong,C.A. *et al.* Genetic, structural and functional diversities of human complement components C4A and C4B and their mouse homologues, Slp and C4. *Int. Immunopharmacol.* **1**, 365-392 (2001).
  124. Truedsson,L., Bengtsson,A.A., & Sturfelt,G. Complement deficiencies and systemic lupus erythematosus. *Autoimmunity* **40**, 560-566 (2007).

125. Pickering,M.C. & Walport,M.J. Links between complement abnormalities and systemic lupus erythematosus. *Rheumatology. (Oxford)* **39**, 133-141 (2000).
126. Yang,Y. *et al.* Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. *Am. J. Hum. Genet.* **80**, 1037-1054 (2007).
127. Namjou,B. *et al.* Evaluation of C1q genomic region in minority racial groups of lupus. *Genes Immun.* **10**, 517-524 (2009).
128. Racila,D.M. *et al.* Homozygous single nucleotide polymorphism of the complement C1QA gene is associated with decreased levels of C1q in patients with subacute cutaneous lupus erythematosus. *Lupus* **12**, 124-132 (2003).
129. Chowdhury,D. *et al.* The exonuclease TREX1 is in the SET complex and acts in concert with NM23-H1 to degrade DNA during granzyme A-mediated cell death. *Mol. Cell* **23**, 133-142 (2006).
130. Stetson,D.B., Ko,J.S., Heidmann,T., & Medzhitov,R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* **134**, 587-598 (2008).
131. Ramantani,G. *et al.* Expanding the phenotypic spectrum of lupus erythematosus in Aicardi-Goutieres syndrome. *Arthritis Rheum.* **62**, 1469-1477 (2010).
132. Lee-Kirsch,M.A. *et al.* Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. *Nat. Genet.* **39**, 1065-1067 (2007).
133. Namjou,B. *et al.* Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort. *Genes Immun.* **12**, 270-279 (2011).
134. Bredius,R.G. *et al.* Phagocytosis of Staphylococcus aureus and Haemophilus influenzae type B opsonized with polyclonal human IgG1 and IgG2 antibodies. Functional hFc gamma RIIa polymorphism to IgG2. *J. Immunol.* **151**, 1463-1472 (1993).
135. Chen,J.Y. *et al.* Fc gamma receptor IIa, IIIa, and IIIb polymorphisms of systemic lupus erythematosus in Taiwan. *Ann. Rheum. Dis.* **63**, 877-880 (2004).
136. Duits,A.J. *et al.* Skewed distribution of IgG Fc receptor IIa (CD32) polymorphism is associated with renal disease in systemic lupus erythematosus patients. *Arthritis Rheum.* **38**, 1832-1836 (1995).
137. Salmon,J.E. *et al.* Fc gamma RIIA alleles are heritable risk factors for lupus nephritis in African Americans. *J. Clin. Invest* **97**, 1348-1354 (1996).
138. Song,Y.W. *et al.* Abnormal distribution of Fc gamma receptor type IIa polymorphisms in Korean patients with systemic lupus erythematosus. *Arthritis Rheum.* **41**, 421-426 (1998).
139. Yap,S.N., Phipps,M.E., Manivasagar,M., Tan,S.Y., & Bosco,J.J. Human Fc gamma receptor IIa (Fc gamma RIIA) genotyping and association with systemic



- lupus erythematosus (SLE) in Chinese and Malays in Malaysia. *Lupus* **8**, 305-310 (1999).
140. Karassa,F.B., Trikalinos,T.A., & Ioannidis,J.P. The Fc gamma RIIIA-F158 allele is a risk factor for the development of lupus nephritis: a meta-analysis. *Kidney Int.* **63**, 1475-1482 (2003).
  141. Koene,H.R. *et al.* The Fc gammaRIIIA-158F allele is a risk factor for systemic lupus erythematosus. *Arthritis Rheum.* **41**, 1813-1818 (1998).
  142. van der Pol,W. & van de Winkel,J.G. IgG receptor polymorphisms: risk factors for disease. *Immunogenetics* **48**, 222-232 (1998).
  143. Alarcon,G.S. *et al.* Time to renal disease and end-stage renal disease in PROFILE: a multiethnic lupus cohort. *PLoS. Med.* **3**, e396 (2006).
  144. Zuniga,R. *et al.* Identification of IgG subclasses and C-reactive protein in lupus nephritis: the relationship between the composition of immune deposits and FCgamma receptor type IIA alleles. *Arthritis Rheum.* **48**, 460-470 (2003).
  145. Sullivan,K.E. *et al.* Analysis of polymorphisms affecting immune complex handling in systemic lupus erythematosus. *Rheumatology. (Oxford)* **42**, 446-452 (2003).
  146. Chu,Z.T. *et al.* Association of Fcgamma receptor IIb polymorphism with susceptibility to systemic lupus erythematosus in Chinese: a common susceptibility gene in the Asian populations. *Tissue Antigens* **63**, 21-27 (2004).
  147. Kyogoku,C. *et al.* Fcgamma receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: contribution of FCGR2B to genetic susceptibility. *Arthritis Rheum.* **46**, 1242-1254 (2002).
  148. Su,K. *et al.* A promoter haplotype of the immunoreceptor tyrosine-based inhibitory motif-bearing FcgammaRIIb alters receptor expression and associates with autoimmunity. II. Differential binding of GATA4 and Yin-Yang1 transcription factors and correlated receptor expression and function. *J. Immunol.* **172**, 7192-7199 (2004).
  149. Kyogoku,C., Tsuchiya,N., Wu,H., Tsao,B.P., & Tokunaga,K. Association of Fcgamma receptor IIA, but not IIB and IIIA, polymorphisms with systemic lupus erythematosus: A family-based association study in Caucasians. *Arthritis Rheum.* **50**, 671-673 (2004).
  150. Li,X. *et al.* A novel polymorphism in the Fcgamma receptor IIB (CD32B) transmembrane region alters receptor signaling. *Arthritis Rheum.* **48**, 3242-3252 (2003).
  151. Magnusson,V. *et al.* Polymorphisms of the Fc gamma receptor type IIB gene are not associated with systemic lupus erythematosus in the Swedish population. *Arthritis Rheum.* **50**, 1348-1350 (2004).
  152. Floto,R.A. *et al.* Loss of function of a lupus-associated FcgammaRIIb polymorphism through exclusion from lipid rafts. *Nat. Med.* **11**, 1056-1058 (2005).

153. Fanciulli,M. *et al.* FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat. Genet.* **39**, 721-723 (2007).
154. Willcocks,L.C. *et al.* Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. *J. Exp. Med.* **205**, 1573-1582 (2008).
155. Sigurdsson,S. *et al.* Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.* **76**, 528-537 (2005).
156. Cunninghame Graham,D.S. *et al.* Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS. Genet.* **7**, e1002341 (2011).
157. Hellquist,A. *et al.* Evidence for genetic association and interaction between the TYK2 and IRF5 genes in systemic lupus erythematosus. *J. Rheumatol.* **36**, 1631-1638 (2009).
158. Suarez-Gestal,M. *et al.* Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. *Arthritis Res. Ther.* **11**, R69 (2009).
159. Budarf,M.L. *et al.* A targeted association study in systemic lupus erythematosus identifies multiple susceptibility alleles. *Genes Immun.* **12**, 51-58 (2011).
160. Cunninghame Graham,D.S., Akil,M., & Vyse,T.J. Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families. *Rheumatology. (Oxford)* **46**, 927-930 (2007).
161. Li,P., Chang,Y.K., Shek,K.W., & Lau,Y.L. Lack of association of TYK2 gene polymorphisms in Chinese patients with systemic lupus erythematosus. *J. Rheumatol.* **38**, 177-178 (2011).
162. Sandling,J.K. *et al.* A candidate gene study of the type I interferon pathway implicates IKBKE and IL8 as risk loci for SLE. *Eur. J. Hum. Genet.* **19**, 479-484 (2011).
163. Jawaheer,D. *et al.* Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum.* **48**, 906-916 (2003).
164. Taylor,K.E. *et al.* Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS. Genet.* **4**, e1000084 (2008).
165. Kawasaki,A. *et al.* Role of STAT4 polymorphisms in systemic lupus erythematosus in a Japanese population: a case-control association study of the STAT1-STAT4 region. *Arthritis Res. Ther.* **10**, R113 (2008).
166. Kobayashi,S. *et al.* Association of STAT4 with susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Japanese population. *Arthritis Rheum.* **58**, 1940-1946 (2008).

167. Yang,W. *et al.* Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXX, are associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.* **10**, 219-226 (2009).
168. Ji,J.D. *et al.* Association of STAT4 polymorphism with rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. *Mol. Biol. Rep.* **37**, 141-147 (2010).
169. Abelson,A.K. *et al.* STAT4 associates with systemic lupus erythematosus through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk. *Ann. Rheum. Dis.* **68**, 1746-1753 (2009).
170. Namjou,B. *et al.* High-density genotyping of STAT4 reveals multiple haplotypic associations with systemic lupus erythematosus in different racial groups. *Arthritis Rheum.* **60**, 1085-1095 (2009).
171. Sigurdsson,S. *et al.* A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. *Hum. Mol. Genet.* **17**, 2868-2876 (2008).
172. Sanchez,E. *et al.* Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus. *Ann. Rheum. Dis.* **70**, 1752-1757 (2011).
173. Svenungsson,E. *et al.* A STAT4 risk allele is associated with ischaemic cerebrovascular events and anti-phospholipid antibodies in systemic lupus erythematosus. *Ann. Rheum. Dis.* **69**, 834-840 (2010).
174. Taylor,K.E. *et al.* Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. *PLoS. Genet.* **7**, e1001311 (2011).
175. Kariuki,S.N. *et al.* Cutting edge: autoimmune disease risk variant of STAT4 confers increased sensitivity to IFN-alpha in lupus patients in vivo. *J. Immunol.* **182**, 34-38 (2009).
176. Finnegan,A. *et al.* IL-4 and IL-12 regulate proteoglycan-induced arthritis through Stat-dependent mechanisms. *J. Immunol.* **169**, 3345-3352 (2002).
177. Watford,W.T. *et al.* Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. *Immunol. Rev.* **202**, 139-156 (2004).
178. Zhang,J. *et al.* The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. *Nat. Genet.* **43**, 902-907 (2011).
179. Chung,S.A. & Criswell,L.A. PTPN22: its role in SLE and autoimmunity. *Autoimmunity* **40**, 582-590 (2007).
180. Eliopoulos,E. *et al.* Association of the PTPN22 R620W polymorphism with increased risk for SLE in the genetically homogeneous population of Crete. *Lupus* **20**, 501-506 (2011).
181. Piotrowski,P., Lianeri,M., Wudarski,M., Lacki,J.K., & Jagodzinski,P.P. Contribution of the R620W polymorphism of protein tyrosine phosphatase non-receptor 22 to systemic lupus erythematosus in Poland. *Clin. Exp. Rheumatol.* **26**, 1099-1102 (2008).

182. Lea,W.W. & Lee,Y.H. The association between the PTPN22 C1858T polymorphism and systemic lupus erythematosus: a meta-analysis update. *Lupus* **20**, 51-57 (2011).
183. Vang,T. *et al.* Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat. Genet.* **37**, 1317-1319 (2005).
184. Rieck,M. *et al.* Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J. Immunol.* **179**, 4704-4710 (2007).
185. Orru,V. *et al.* A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. *Hum. Mol. Genet.* **18**, 569-579 (2009).
186. Prokunina,L. *et al.* A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. *Nat. Genet.* **32**, 666-669 (2002).
187. Ferreiros-Vidal,I. *et al.* Bias in association studies of systemic lupus erythematosus susceptibility due to geographical variation in the frequency of a programmed cell death 1 polymorphism across Europe. *Genes Immun.* **8**, 138-146 (2007).
188. Beyaert,R., Heyninck,K., & Van,H.S. A20 and A20-binding proteins as cellular inhibitors of nuclear factor-kappa B-dependent gene expression and apoptosis. *Biochem. Pharmacol.* **60**, 1143-1151 (2000).
189. Boone,D.L. *et al.* The ubiquitin-modifying enzyme A20 is required for termination of Toll-like receptor responses. *Nat. Immunol.* **5**, 1052-1060 (2004).
190. Wertz,I.E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature* **430**, 694-699 (2004).
191. Musone,S.L. *et al.* Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 1062-1064 (2008).
192. Bates,J.S. *et al.* Meta-analysis and imputation identifies a 109 kb risk haplotype spanning TNFAIP3 associated with lupus nephritis and hematologic manifestations. *Genes Immun.* **10**, 470-477 (2009).
193. Adrianto,I. *et al.* Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. *Nat. Genet.* **43**, 253-258 (2011).
194. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).
195. Dieguez-Gonzalez,R. *et al.* Analysis of TNFAIP3, a feedback inhibitor of nuclear factor-kappaB and the neighbor intergenic 6q23 region in rheumatoid arthritis susceptibility. *Arthritis Res. Ther.* **11**, R42 (2009).
196. Thomson,W. *et al.* Rheumatoid arthritis association at 6q23. *Nat. Genet.* **39**, 1431-1433 (2007).

197. Musone,S.L. *et al.* Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 1062-1064 (2008).
198. Heyninck,K., Kreike,M.M., & Beyaert,R. Structure-function analysis of the A20-binding inhibitor of NF-kappa B activation, ABIN-1. *FEBS Lett.* **536**, 135-140 (2003).
199. Coxon,A. *et al.* A novel role for the beta 2 integrin CD11b/CD18 in neutrophil apoptosis: a homeostatic mechanism in inflammation. *Immunity.* **5**, 653-666 (1996).
200. Dunne,J.L., Collins,R.G., Beaudet,A.L., Ballantyne,C.M., & Ley,K. Mac-1, but not LFA-1, uses intercellular adhesion molecule-1 to mediate slow leukocyte rolling in TNF-alpha-induced inflammation. *J. Immunol.* **171**, 6105-6111 (2003).
201. Shappell,S.B. *et al.* Mac-1 (CD11b/CD18) mediates adherence-dependent hydrogen peroxide production by human and canine neutrophils. *J. Immunol.* **144**, 2702-2711 (1990).
202. Nath,S.K. *et al.* A nonsynonymous functional variant in integrin-alpha(M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 152-154 (2008).
203. Han,S. *et al.* Evaluation of imputation-based association in and around the integrin-alpha-M (ITGAM) gene and replication of robust association between a non-synonymous functional variant within ITGAM and systemic lupus erythematosus (SLE). *Hum. Mol. Genet.* **18**, 1171-1180 (2009).
204. Yang,W. *et al.* ITGAM is associated with disease susceptibility and renal nephritis of systemic lupus erythematosus in Hong Kong Chinese and Thai. *Hum. Mol. Genet.* **18**, 2063-2070 (2009).
205. Tretter,T., Ross,A.E., Dordai,D.I., & Desiderio,S. Mimicry of pre-B cell receptor signaling by activation of the tyrosine kinase Blk. *J. Exp. Med.* **198**, 1863-1873 (2003).
206. Wasserman,R., Li,Y.S., & Hardy,R.R. Differential expression of the blk and ret tyrosine kinases during B lineage development is dependent on Ig rearrangement. *J. Immunol.* **155**, 644-651 (1995).
207. Gourh,P. *et al.* Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. *J. Autoimmun.* **34**, 155-162 (2010).
208. Castillejo-Lopez,C. *et al.* Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. *Ann. Rheum. Dis.* **71**, 136-142 (2012).
209. Jarvinen,T.M. *et al.* Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population. *Rheumatology. (Oxford)* **51**, 87-92 (2012).
210. Zhou,X.J. *et al.* Gene-gene interaction of BLK, TNFSF4, TRAF1, TNFAIP3, and REL in systemic lupus erythematosus. *Arthritis Rheum.* **64**, 222-231 (2012).

- 
211. Yokoyama,K. *et al.* BANK regulates BCR-induced calcium mobilization by promoting tyrosine phosphorylation of IP(3) receptor. *EMBO J.* **21**, 83-92 (2002).
212. Lu,R. *et al.* Genetic associations of LYN with systemic lupus erythematosus. *Genes Immun.* **10**, 397-403 (2009).
213. Chahrour,M. *et al.* MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320**, 1224-1229 (2008).
214. Jones,P.L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187-191 (1998).
215. Webb,R. *et al.* Variants within MECP2, a key transcription regulator, are associated with increased susceptibility to lupus and differential gene expression in patients with systemic lupus erythematosus. *Arthritis Rheum.* **60**, 1076-1084 (2009).
216. Baechler,E.C., Gregersen,P.K., & Behrens,T.W. The emerging role of interferon in human systemic lupus erythematosus. *Curr. Opin. Immunol.* **16**, 801-807 (2004).
217. Jacob,C.O., van der Meide,P.H., & McDevitt,H.O. In vivo treatment of (NZB X NZW)F1 lupus-like nephritis with monoclonal antibody to gamma interferon. *J. Exp. Med.* **166**, 798-803 (1987).
218. Wong,H.K., Kammer,G.M., Dennis,G., & Tsokos,G.C. Abnormal NF-kappa B activity in T lymphocytes from patients with systemic lupus erythematosus is associated with decreased p65-RelA protein expression. *J. Immunol.* **163**, 1682-1689 (1999).
219. Jacob,C.O. *et al.* Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U. S. A* **106**, 6256-6261 (2009).
220. Sawalha,A.H. *et al.* Common variants within MECP2 confer risk of systemic lupus erythematosus. *PLoS. One.* **3**, e1727 (2008).
221. Webb,R. *et al.* Variants within MECP2, a key transcription regulator, are associated with increased susceptibility to lupus and differential gene expression in patients with systemic lupus erythematosus. *Arthritis Rheum.* **60**, 1076-1084 (2009).
222. Jacob,C.O. *et al.* Identification of novel susceptibility genes in childhood-onset systemic lupus erythematosus using a uniquely designed candidate gene pathway platform. *Arthritis Rheum.* **56**, 4164-4173 (2007).
223. Jacob,C.O. *et al.* Identification of IRAK1 as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U. S. A* **106**, 6256-6261 (2009).
224. Sawalha,A.H. *et al.* Common variants within MECP2 confer risk of systemic lupus erythematosus. *PLoS. One.* **3**, e1727 (2008).
225. Ito,T. *et al.* OX40 ligand shuts down IL-10-producing regulatory T cells. *Proc. Natl. Acad. Sci. U. S. A* **103**, 13138-13143 (2006).

- 
226. Stuber,E., Neurath,M., Calderhead,D., Fell,H.P., & Strober,W. Cross-linking of OX40 ligand, a member of the TNF/NGF cytokine family, induces proliferation and differentiation in murine splenic B cells. *Immunity*. **2**, 507-521 (1995).
227. Delgado-Vega,A.M. *et al.* Replication of the TNFSF4 (OX40L) promoter region association with systemic lupus erythematosus. *Genes Immun.* **10**, 248-253 (2009).
228. Cunninghame Graham,D.S. *et al.* Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nat. Genet.* **40**, 83-89 (2008).
229. Salloom,R. *et al.* Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients. *Arthritis Rheum.* **62**, 553-561 (2010).
230. Crotty,S., Johnston,R.J., & Schoenberger,S.P. Effectors and memories: Bcl-6 and Blimp-1 in T and B lymphocyte differentiation. *Nat. Immunol.* **11**, 114-120 (2010).
231. Wang,D. *et al.* Ets-1 deficiency leads to altered B cell differentiation, hyperresponsiveness to TLR9 and autoimmune disease. *Int. Immunol.* **17**, 1179-1191 (2005).
232. Cunninghame Graham,D.S. *et al.* Association of LY9 in UK and Canadian SLE families. *Genes Immun.* **9**, 93-102 (2008).
233. Zhang,Y. *et al.* Two missense variants in UHRF1BP1 are independently associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.* **12**, 231-234 (2011).
234. Agik,S. *et al.* The Autoimmune Disease Risk Allele of UBE2L3 in African American Patients with Systemic Lupus Erythematosus: A Recessive Effect Upon Subphenotypes. *J. Rheumatol.* **39**, 73-78 (2012).
235. Chung,S.A. *et al.* Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS. Genet.* **7**, e1001323 (2011).
236. Diveu,C., McGeachy,M.J., & Cua,D.J. Cytokines that regulate autoimmunity. *Curr. Opin. Immunol.* **20**, 663-668 (2008).
237. Hagiwara,E., Gourley,M.F., Lee,S., & Klinman,D.K. Disease severity in patients with systemic lupus erythematosus correlates with an increased ratio of interleukin-10:interferon-gamma-secreting cells in the peripheral blood. *Arthritis Rheum.* **39**, 379-385 (1996).
238. Nath,S.K., Harley,J.B., & Lee,Y.H. Polymorphisms of complement receptor 1 and interleukin-10 genes and systemic lupus erythematosus: a meta-analysis. *Hum. Genet.* **118**, 225-234 (2005).
239. Kyogoku,C. & Tsuchiya,N. A compass that points to lupus: genetic studies on type I interferon pathway. *Genes Immun.* **8**, 445-455 (2007).
-

- 
240. Izaguirre,A. *et al.* Comparative analysis of IRF and IFN-alpha expression in human plasmacytoid and monocyte-derived dendritic cells. *J. Leukoc. Biol.* **74**, 1125-1138 (2003).
241. Mancl,M.E. *et al.* Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J. Biol. Chem.* **280**, 21078-21090 (2005).
242. Taniguchi,T., Ogasawara,K., Takaoka,A., & Tanaka,N. IRF family of transcription factors as regulators of host defense. *Annu. Rev. Immunol.* **19**, 623-655 (2001).
243. Graham,R.R. *et al.* A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.* **38**, 550-555 (2006).
244. Honda,K. & Taniguchi,T. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.* **6**, 644-658 (2006).
245. Krausgruber,T. *et al.* IRF5 promotes inflammatory macrophage polarization and TH1-TH17 responses. *Nat. Immunol.* **12**, 231-238 (2011).
246. Barnes,B.J., Kellum,M.J., Field,A.E., & Pitha,P.M. Multiple regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes. *Mol. Cell Biol.* **22**, 5721-5740 (2002).
247. Barnes,B.J., Field,A.E., & Pitha-Rowe,P.M. Virus-induced heterodimer formation between IRF-5 and IRF-7 modulates assembly of the IFNA enhanceosome in vivo and transcriptional activity of IFNA genes. *J. Biol. Chem.* **278**, 16630-16641 (2003).
248. Mori,T. *et al.* Identification of the interferon regulatory factor 5 gene (IRF-5) as a direct target for p53. *Oncogene* **21**, 2914-2918 (2002).
249. Savitsky,D., Tamura,T., Yanai,H., & Taniguchi,T. Regulation of immunity and oncogenesis by the IRF transcription factor family. *Cancer Immunol. Immunother.* **59**, 489-510 (2010).
250. Barnes,B.J., Kellum,M.J., Pinder,K.E., Frisancho,J.A., & Pitha,P.M. Interferon regulatory factor 5, a novel mediator of cell cycle arrest and cell death. *Cancer Res.* **63**, 6424-6431 (2003).
251. Hu,G., Mancl,M.E., & Barnes,B.J. Signaling through IFN regulatory factor-5 sensitizes p53-deficient tumors to DNA damage-induced apoptosis and cell death. *Cancer Res.* **65**, 7403-7412 (2005).
252. Yanai,H. *et al.* Role of IFN regulatory factor 5 transcription factor in antiviral immunity and tumor suppression. *Proc. Natl. Acad. Sci. U. S. A* **104**, 3402-3407 (2007).
253. Paun,A. *et al.* Functional characterization of murine interferon regulatory factor 5 (IRF-5) and its role in the innate antiviral response. *J. Biol. Chem.* **283**, 14295-14308 (2008).
-



- 
254. Takaoka,A. *et al.* Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* **434**, 243-249 (2005).
255. Tada,Y. *et al.* Interferon regulatory factor 5 is critical for the development of lupus in MRL/lpr mice. *Arthritis Rheum.* **63**, 738-748 (2011).
256. Cohen,P.L. & Eisenberg,R.A. Lpr and gld: single gene models of systemic autoimmunity and lymphoproliferative disease. *Annu. Rev. Immunol.* **9**, 243-269 (1991).
257. Watanabe-Fukunaga,R., Brannan,C.I., Copeland,N.G., Jenkins,N.A., & Nagata,S. Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis. *Nature* **356**, 314-317 (1992).
258. Watson,M.L. *et al.* Genetic analysis of MRL-lpr mice: relationship of the Fas apoptosis gene to disease manifestations and renal disease-modifying loci. *J. Exp. Med.* **176**, 1645-1656 (1992).
259. Bave,U. *et al.* Fc gamma RIIa is expressed on natural IFN-alpha-producing cells (plasmacytoid dendritic cells) and is required for the IFN-alpha production induced by apoptotic cells combined with lupus IgG. *J. Immunol.* **171**, 3296-3302 (2003).
260. Balkhi,M.Y., Fitzgerald,K.A., & Pitha,P.M. Functional regulation of MyD88-activated interferon regulatory factor 5 by K63-linked polyubiquitination. *Mol. Cell Biol.* **28**, 7296-7308 (2008).
261. Honda,K., Yanai,H., Takaoka,A., & Taniguchi,T. Regulation of the type I IFN induction: a current view. *Int. Immunol.* **17**, 1367-1378 (2005).
262. Akahoshi,M., Nakashima,H., & Shirakawa,T. Roles of genetic variations in signalling/immunoregulatory molecules in susceptibility to systemic lupus erythematosus. *Semin. Immunol.* **18**, 224-229 (2006).
263. Balkhi,M.Y., Fitzgerald,K.A., & Pitha,P.M. Functional regulation of MyD88-activated interferon regulatory factor 5 by K63-linked polyubiquitination. *Mol. Cell Biol.* **28**, 7296-7308 (2008).
264. Barnes,B.J. *et al.* Global and distinct targets of IRF-5 and IRF-7 during innate response to viral infection. *J. Biol. Chem.* **279**, 45194-45207 (2004).
265. Kelly,J.A. *et al.* Interferon regulatory factor-5 is genetically associated with systemic lupus erythematosus in African Americans. *Genes Immun.* **9**, 187-194 (2008).
266. Niewold,T.B. *et al.* IRF5 haplotypes demonstrate diverse serological associations which predict serum interferon alpha activity and explain the majority of the genetic association with systemic lupus erythematosus. *Ann. Rheum. Dis.*(2011).
267. Hellquist,A. *et al.* Evidence for genetic association and interaction between the TYK2 and IRF5 genes in systemic lupus erythematosus. *J. Rheumatol.* **36**, 1631-1638 (2009).
268. Lofgren,S.E. *et al.* Promoter insertion/deletion in the IRF5 gene is highly associated with susceptibility to systemic lupus erythematosus in distinct

- populations, but exerts a modest effect on gene expression in peripheral blood mononuclear cells. *J. Rheumatol.* **37**, 574-578 (2010).
269. Hu,W. & Ren,H. A meta-analysis of the association of IRF5 polymorphism with systemic lupus erythematosus. *Int. J. Immunogenet.* **38**, 411-417 (2011).
  270. Lee,Y.H. & Song,G.G. Association between the rs2004640 functional polymorphism of interferon regulatory factor 5 and systemic lupus erythematosus: a meta-analysis. *Rheumatol. Int.* **29**, 1137-1142 (2009).
  271. Cunninghame Graham,D.S. *et al.* Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. *Hum. Mol. Genet.* **16**, 579-591 (2007).
  272. Feng,D. *et al.* Genetic variants and disease-associated factors contribute to enhanced interferon regulatory factor 5 expression in blood cells of patients with systemic lupus erythematosus. *Arthritis Rheum.* **62**, 562-573 (2010).
  273. Graham,R.R. *et al.* Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. *Proc. Natl. Acad. Sci. U. S. A* **104**, 6758-6763 (2007).
  274. Kozyrev,S.V. *et al.* Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus. *Arthritis Rheum.* **56**, 1234-1241 (2007).
  275. Niewold,T.B. *et al.* Association of the IRF5 risk haplotype with high serum interferon-alpha activity in systemic lupus erythematosus patients. *Arthritis Rheum.* **58**, 2481-2487 (2008).
  276. Demirci,F.Y. *et al.* Association of a common interferon regulatory factor 5 (IRF5) variant with increased risk of systemic lupus erythematosus (SLE). *Ann. Hum. Genet.* **71**, 308-311 (2007).
  277. Nordang,G.B. *et al.* Interferon regulatory factor 5 gene polymorphism confers risk to several rheumatic diseases and correlates with expression of alternative thymic transcripts. *Rheumatology. (Oxford)*(2011).
  278. Dieguez-Gonzalez,R. *et al.* Association of interferon regulatory factor 5 haplotypes, similar to that found in systemic lupus erythematosus, in a large subgroup of patients with rheumatoid arthritis. *Arthritis Rheum.* **58**, 1264-1274 (2008).
  279. Cheung,V.G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365-1369 (2005).
  280. Morley,M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743-747 (2004).
  281. Wen,F., Ellingson,S.M., Kyogoku,C., Peterson,E.J., & Gaffney,P.M. Exon 6 variants carried on systemic lupus erythematosus (SLE) risk haplotypes modulate IRF5 function. *Autoimmunity* **44**, 82-89 (2011).

- 
282. Dideberg,V. *et al.* An insertion-deletion polymorphism in the interferon regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. *Hum. Mol. Genet.* **16**, 3008-3016 (2007).
283. Sigurdsson,S. *et al.* Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (IRF5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus. *Hum. Mol. Genet.* **17**, 872-881 (2008).
284. Wang,Z., Gerstein,M., & Snyder,M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63 (2009).
285. Alarcon,G.S. *et al.* Systemic lupus erythematosus in three ethnic groups: III. A comparison of characteristics early in the natural history of the LUMINA cohort. LUPus in MInority populations: NAture vs. Nurture. *Lupus* **8**, 197-209 (1999).
286. Alarcon,G.S. *et al.* Baseline characteristics of a multiethnic lupus cohort: PROFILE. *Lupus* **11**, 95-101 (2002).
287. Bossingham,D. Systemic lupus erythematosus in the far north of Queensland. *Lupus* **12**, 327-331 (2003).
288. Peschken,C.A. *et al.* The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. *J. Rheumatol.* **36**, 1200-1208 (2009).
289. Molina,J.F. *et al.* Ethnic differences in the clinical expression of systemic lupus erythematosus: a comparative study between African-Americans and Latin Americans. *Lupus* **6**, 63-67 (1997).
290. Pons-Estel,B.A. *et al.* The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics". *Medicine (Baltimore)* **83**, 1-17 (2004).
291. Borchers,A.T., Naguwa,S.M., Shoenfeld,Y., & Gershwin,M.E. The geoepidemiology of systemic lupus erythematosus. *Autoimmun. Rev.* **9**, A277-A287 (2010).
292. Seldin,M.F. *et al.* European population substructure: clustering of northern and southern populations. *PLoS. Genet.* **2**, e143 (2006).
293. Han,J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS. Genet.* **4**, e1000074 (2008).
294. Tian,C. *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS. Genet.* **4**, e4 (2008).
295. Tian,C. *et al.* European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Mol. Med.* **15**, 371-383 (2009).
296. Chung,S.A. *et al.* European population substructure is associated with mucocutaneous manifestations and autoantibody production in systemic lupus erythematosus. *Arthritis Rheum.* **60**, 2448-2456 (2009).

- 
297. Richman,I.B. *et al.* European population substructure correlates with systemic lupus erythematosus endophenotypes in North Americans of European descent. *Genes Immun.* **11**, 515-521 (2010).
298. Sanchez,E. *et al.* Genetically determined Amerindian ancestry correlates with increased frequency of risk alleles for systemic lupus erythematosus. *Arthritis Rheum.* **62**, 3722-3729 (2010).
299. Kim-Howard,X. *et al.* ITGAM coding variant (rs1143679) influences the risk of renal disease, discoid rash and immunological manifestations in patients with systemic lupus erythematosus with European ancestry. *Ann. Rheum. Dis.* **69**, 1329-1332 (2010).
300. Barrett,J.C., Fry,B., Maller,J., & Daly,M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* **21**, 263-265 (2005).
301. Stephens,M., Smith,N.J., & Donnelly,P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978-989 (2001).
302. Miceli-Richard,C. *et al.* Association of an IRF5 gene functional polymorphism with Sjogren's syndrome. *Arthritis Rheum.* **56**, 3989-3994 (2007).
303. Miceli-Richard,C. *et al.* The CGGGG insertion/deletion polymorphism of the IRF5 promoter is a strong risk factor for primary Sjogren's syndrome. *Arthritis Rheum.* **60**, 1991-1997 (2009).
304. Dieude,P. *et al.* Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. *Arthritis Rheum.* **60**, 225-233 (2009).
305. Dieude,P. *et al.* Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. *J. Rheumatol.* **37**, 987-992 (2010).
306. Hirschfield,G.M. *et al.* Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis. *Nat. Genet.* **42**, 655-657 (2010).
307. Liu,X. *et al.* Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nat. Genet.* **42**, 658-660 (2010).
308. Dixon,A.L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39**, 1202-1207 (2007).
309. Kwan,T. *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.* **40**, 225-231 (2008).
310. Stranger,B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217-1224 (2007).
311. Kristjansdottir,G. *et al.* Interferon regulatory factor 5 (IRF5) gene variants are associated with multiple sclerosis in three distinct populations. *J. Med. Genet.* **45**, 362-369 (2008).
312. Montgomery,S.B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777 (2010).
-

- 
313. Bauchet,M. *et al.* Measuring European population stratification with microarray genotype data. *Am. J. Hum. Genet.* **80**, 948-956 (2007).
314. Bersaglieri,T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111-1120 (2004).
315. Arnett,F.C., Reveille,J.D., Moutsopoulos,H.M., Georgescu,L., & Elkon,K.B. Ribosomal P autoantibodies in systemic lupus erythematosus. Frequencies in different ethnic groups and clinical and immunogenetic associations. *Arthritis Rheum.* **39**, 1833-1839 (1996).
316. Garcia,C.O. *et al.* Autoantibody profile in African-American patients with lupus nephritis. *Lupus* **5**, 602-605 (1996).
317. Sestak,A.L. *et al.* Patients with familial and sporadic onset SLE have similar clinical profiles but vary profoundly by race. *Lupus* **17**, 1004-1009 (2008).
318. Tikly,M., Burgin,S., Mohanlal,P., Bellingan,A., & George,J. Autoantibodies in black South Africans with systemic lupus erythematosus: spectrum and clinical associations. *Clin. Rheumatol.* **15**, 261-265 (1996).
319. To,C.H. & Petri,M. Is antibody clustering predictive of clinical subsets and damage in systemic lupus erythematosus?. *Arthritis Rheum.* **52**, 4003-4010 (2005).
320. Vazquez-Del,M.M. *et al.* High prevalence of autoantibodies to RNA helicase A in Mexican patients with systemic lupus erythematosus. *Arthritis Res. Ther.* **12**, R6 (2010).
321. Yamasaki,Y. *et al.* Autoantibodies to RNA helicase A: a new serologic marker of early lupus. *Arthritis Rheum.* **56**, 596-604 (2007).
322. Lahita,R. Systemic Lupus Erythematosus. 4th ed. 4th ed., 109-151. 2004. San Diego, Elsevier Academic Press. MHC Class II and non-MHC genes in the pathogenesis of systemic lupus erythematosus. Ref Type: Edited Book
323. Podrebarac,T.A., Boisert,D.M., & Goldstein,R. Clinical correlates, serum autoantibodies and the role of the major histocompatibility complex in French Canadian and non-French Canadian Caucasians with SLE. *Lupus* **7**, 183-191 (1998).
324. Hartung,K. *et al.* MHC gene products and anticardiolipin antibodies in systemic lupus erythematosus results of a multicenter study. SLE Study Group. *Autoimmunity* **13**, 95-99 (1992).
325. Sanchez,M.L. *et al.* Association of HLA-DM polymorphism with the production of antiphospholipid antibodies. *Ann. Rheum. Dis.* **63**, 1645-1648 (2004).
326. (2006). [Registrolesaf. Features of patients with systemic lupus erythematosus and antiphospholipid syndrome registered by rheumatologists and internists]. 131-136. 2012. Ref Type: Edited Book
327. Santos,M.J. *et al.* [Characterization of a Portuguese population with systemic lupus erytematosus]. *Acta Reumatol. Port.* **32**, 153-161 (2007).

- 
328. Cervera,R. *et al.* Systemic lupus erythematosus in Europe at the change of the millennium: lessons from the "Euro-Lupus Project". *Autoimmun. Rev.* **5**, 180-186 (2006).
329. Del,B.S., Sok,J., Bessac,E., & Bernerd,F. Relationship between skin response to ultraviolet exposure and skin color type. *Pigment Cell Res.* **19**, 606-614 (2006).
330. Uter,W., Pfahlberg,A., Kalina,B., Kolmel,K.F., & Gefeller,O. Inter-relation between variables determining constitutional UV sensitivity in Caucasian children. *Photodermatol. Photoimmunol. Photomed.* **20**, 9-13 (2004).
331. Gregersen,P.K., Lee,H.S., Batliwalla,F., & Begovich,A.B. PTPN22: setting thresholds for autoimmunity. *Semin. Immunol.* **18**, 214-223 (2006).
332. Cummings M. **Human Heredity: Principles and Issues**. 2012. Brooks/Cole. 2010. Ref Type: Edited Book.
333. Barreiro,L.B. & Quintana-Murci,L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* **11**, 17-30 (2010).
334. Fumagalli,M. *et al.* Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS. Genet.* **7**, e1002355 (2011).
335. Kawasaki,A. *et al.* Association of IRF5 polymorphisms with systemic lupus erythematosus in a Japanese population: support for a crucial role of intron 1 polymorphisms. *Arthritis Rheum.* **58**, 826-834 (2008).
336. Siu,H.O. *et al.* Association of a haplotype of IRF5 gene with systemic lupus erythematosus in Chinese. *J. Rheumatol.* **35**, 360-362 (2008).
337. Rullo,O.J. *et al.* Association of IRF5 polymorphisms with activation of the interferon alpha pathway. *Ann. Rheum. Dis.* **69**, 611-617 (2010).